

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITE DE SAÏDA - Dr MOULAY TAHAR



Faculté des Sciences
Département de Chimie

MEMOIRE

Présenté par :
Ajradi Abderrazeq
En vue de l'obtention du

Diplôme de Master en Chimie

Spécialité : Chimie Théorique et Computationnelle

Thème

Étude QSAR de l'inhibition de la
transcriptase inverse des dérivés HEPT

Soutenu le 26/09/2021 devant le jury composé de :

Présidente
Encadrante
Examinateur

Dr Djallila Missaoui
Dr Rekia KADARI
Dr Ali Rahmouni

MCB
MCB
Professeur

Université de Saida
Université de Saida
Université de Saida

REMERCIEMENTS

Je tiens à remercier toutes les personnes qui ont contribué au succès de mon stage et qui m'ont aidée lors de la rédaction de ce mémoire.

Je voudrais dans un premier temps Je tiens à remercier mon encadreur Reġia KADARI, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter ma réflexion.

Je tiens également à remercier tout particulièrement le professeur Rahmoni Ali pour son soutien et ses conseils tout au long de la réalisation de mon travail.

Je n'oublie pas mes remerciements au professeur Missaoui Djallila pour l'honneur de présider le jury de ce travail.

Je remercie également toute l'équipe pédagogique de l'université de Saida et les intervenants professionnels responsables de ma formation, Tout au long de mon parcours universitaire.

Dédicace

La locomotive de recherche a franchi de nombreux obstacles, cependant, j'ai essayé de la surmonter régulièrement, grâce à Dieu et à Lui. Je dédie ce travail ;

À mes parents, mes frères, ma femme, mes fils et mes amis, ils étaient comme un soutien et un soutien afin de mener à bien la recherche.

Je ne devrais pas oublier mes professeurs qui ont eu le plus grand rôle en me soutenant et en me fournissant des informations précieuses

Prier Dieu Tout-Puissant pour prolonger votre vie et vous bénir avec des générosités.

ملخص

تثبيط إنزيم النسخ العكسي هو جزيء نشط ضد فيروس نقص المناعة البشرية. يمنع الإنزيم الذي يسمح بتوليف الحمض النووي التكميلي من الحمض النووي الريبي الفيروسي. مبدأهم هو التحول إلى شكل نشط في الجسم والاندماج في خيط الحمض النووي الذي يتم تشكيله.

أجريت دراسات QSAR على 107 مشتقات HEPT تم استخدام إجراءين للتحليل الإحصائي وهما بشكل أساسي: الانحدار متعدد الخطوط (MLR) والإدراك متعدد الطبقات (MLP) لتطوير العلاقات بين الواصفات الجزيئية والنشاط البيولوجي لمشتقات HEPT .

تشير نتائجنا إلى نماذج QSAR بناءً على الواصفات التالية: إمكانات التأين، وتقارب الإلكترون، والنعومة، ومؤشر الإلكترون فيليك، ومؤشرات Fukui .

تم تقدير التنبؤ بالنموذج من خلال التحقق الداخلي والخارجي. ولوحظ وجود علاقة ارتباط جيدة بين قيم النشاط التجريبي والنشاط المتوقع، مما يدل على التحقق من الصحة والجودة الجيدة لنماذج QSAR الناتجة.

الكلمات المفتاحية : HEPT، HIV، QSAR، MLP، MLR

Résumé

L'inhibiteur de la transcriptase inverse est une molécule active contre le VIH. Elle bloque l'enzyme qui permet la synthèse d'ADN complémentaire à partir d'ARN viral. Leur principe est de se transformer en une forme active dans l'organisme et de s'intégrer au brin d'ADN en formation.

Des études QSAR ont été effectuées sur 107 dérivés de HEPT. Deux procédures d'analyse statistiques ont été utilisées qui sont principalement : la régression multilinéaire (MLR) et le perceptron multicouches (MLP) pour élaborer les relations entre les descripteurs moléculaires et l'activité biologique des dérivés de HEPT. Nos résultats suggèrent des modèles QSAR en fonction des descripteurs suivants : le potentiel d'ionisation, l'affinité électronique, la mollesse, l'indice d'électrophilicité et indices de Fukui.

La force du modèle a été estimée par la validation interne et externe. Une bonne corrélation entre les valeurs de l'activité expérimentale et prédite a été observée, indiquant la validation et la bonne qualité des modèles QSAR issus.

Mots clés : HEPT, VIH, QSAR, MLR, MLP.

Abstract

Inhibition of reverse transcriptase is an active molecule against HIV. It blocks the enzyme that allows the synthesis of complementary DNA from viral RNA. Their principle is to transform into an active form in the body and to integrate into the DNA strand being formed.

QSAR studies were performed on 107 HEPT derivatives. Two statistical analysis procedures were used which are mainly: multilinear regression (MLR) and multilayer perceptron (MLP) to develop the relationships between molecular descriptors and the biological activity of HEPT derivatives. Our results suggest QSAR models based on the following descriptors: ionization potential, electron affinity, softness, electrophilic index and Fukui indices.

The predictive of the model was estimated by internal and external validation. A good correlation between the values of the experimental and predicted activity was observed, indicating the validation and the good quality of the resulting QSAR models.

Keywords : HEPT, HIV, QSAR, MLR, MLP

Table des matières

ملخص

Résumé

Abstract

Liste des figures

Liste des tableaux

Liste des abréviations

Introduction Générale

Introduction Générale.....	01
Bibliographie.....	02

Chapitre 1 : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH et méthodes de calculs

I.1 Introduction.....	05
I.2 Les inhibiteurs de la transcriptase inverse.....	05
I.3 Aspect structural des dérivés de HEPT.....	05
I.4 Activité anti-VIH des dérivés de HEPT.....	10
I.5 Les approches théoriques	11
I.5.1 Equation de Schrödinger.....	11
I.5.2 Approximation de Born-Oppenheimer (BO).....	11
I.5.3 Approximation Spin-orbitales.....	12
I.5.4 Approximation Hartree-Fock.....	13
I.5.5 La théorie de la fonctionnelle de la densité (DFT).....	13
I.5.5.1 Fondement de la théorie DFT.....	14
I.5.5.2 Approximation locale de la densité (Local Density Approximation, LDA).....	14
I.5.5.3 Approximation du Gradient Généralisé(GGA).....	14
I.5.5.4 Fonctionnelle hybride B3LYP.....	16
I.6 Calcul des descripteurs quantiques.....	16
I.6.1 Les indices de Fukui.....	16

I.6.2 Potentiel d'ionisation (D1).....	17
I.6.3 Affinité électronique (D2).....	17
I.6.4 La dureté chimique et la mollesse globale (D3).	17
I.6.5 Indice d'électrophilicité global (D4).....	18
I.7 Conclusion.....	18
Bibliographie	19

Chapitre 2 : Relations Quantitatives Structures Activités/Propriétés (QSAR/QSPR)

II. 1 Introduction.....	22
II. 2 Modélisation QSAR/QSPR	22
II. 3 Définition et formalisme.....	23
II. 4 Objectifs du QSAR.....	23
II.5 Les étapes de développement des modèles QSAR.....	23
II.5.1 La préparation des données.....	23
II.5.2 Le traitement des données.....	25
II. 5.3 L'évaluation et la validation des résultats.....	27
II.6 Les métriques d'évaluation.....	27
II.7 Les applications du QSAR.....	28
II.8 Conclusion	29
Bibliographie	30

Chapitre 3 : Résultats et Discussions

III.1 Introduction.....	33
III.2 Préparation et prétraitement des données.....	33
III.3 Sélection des données d'apprentissage et de test.....	41
III.4 Etude quantitative des relations structure-activité QSAR.....	45
III.5 Discussion.....	55
III.6 Conclusion.....	56
Bibliographie	57

Conclusion Générale

Conclusion générale.....	60
--------------------------	----

Liste des figures

Figure I. 1 Numérotation des différents atomes.....	06
Figure I. 2 1-[(2-hydroxyéthoxy) méthyl] -6- (phénylthio) thymine.....	09
Figure I. 3 HEPT 06.....	09
Figure I. 4 HEPT 28.....	09
Figure I. 5 HEPT 106.....	09
Figure I. 6 HEPT 36.....	10
Figure III. 1 Perceptron multicouche du modèle développé pour les dérivés HEPT.....	52

Liste des Tableaux

Tableau I.1 L'ensemble des radicaux dans les différents dérivés de HEPT	06
Tableau III.1 Les valeurs des descripteurs.....	34
Tableau III.2 Structures chimiques et activités anti-VIH observées des dérivés HEPT.....	37
Tableau III.3 Les données d'apprentissage (training).....	41
Tableau III.4 Les données de test.....	45
Tableau III.5 Comparaison de différentes méthodes d'analyse QSAR de dérivés HEPT.....	46
Tableau III.6 Performance du MLP.....	47
Tableau III.7 Les valeurs des activités observées/calculées en utilisant le MLP.....	47
Tableau III.8 Performance de la régression linéaire sur les données d'apprentissage et les données de test pour les dérivés HEPT.....	49
Tableau III.9 Les valeurs expérimentales, prédites et résiduelles des activités par MLR sur les données d'apprentissage.....	49
Tableau III.10 Les valeurs expérimentales, prédites et résiduelles des activités par MLR sur les données de test.....	51
Tableau III.11 Performance du perceptron multi couches sur les données d'apprentissage et les données de test pour les dérivés HEPT.....	52
Tableau III.12 Les valeurs expérimentales, prédites et résiduelles des activités par MLP sur les données d'apprentissage.....	52
Tableau III.13 Les valeurs expérimentales, prédites et résiduelles des activités par MLP sur les données de test.....	55

Liste Des Abréviations

VIH	Virus de l'immunodéficience humaine
ADN	L'acide DésoxyriboNucléique
ARN	Acide RiboNucléique
HEPT	1 - [(2-hydroxyéthoxy) méthyl] -6- (phénylthio) thymine
B3LYP	Becke, Three-Parameter, Lee-Yang-Parr
DFT	Density Functional Theory
BO	Approximation de Born-Oppenheimer
MQ	Mécanique Quantique
LDA	Local Density Approximation
GGA	Approximation du Gradient Généralisé
QSAR	Quantitative Structure activity Relationship
QSPR	Quantitative Structure Property Relationship
RNA	Réseaux de Neurones Artificielle
ACP	L'analyse en composantes principales
MLP	Perceptron Multi Linéaire
MLR	Régression Multi Linéaire

Introduction générale

Introduction générale

En juin 1981, des scientifiques américains ont signalé le premier cas de virus de l'immunodéficience humaine VIH qui est une espèce de rétrovirus infectant l'humain et responsable du syndrome d'immunodéficience acquise rendant système immunitaire vulnérable à de multiples infections opportunistes [1].

La transmission du VIH se fait par contact avec des liquides biologiques (sang, sécrétions vaginales, sperme ou lait maternel) [2] corporels qui contient le virus ou des cellules infectées par le virus.

Véritablement, il n'existe aucun remède pour éliminer complètement le VIH dans le corps. Un traitement approprié peut permettre aux personnes infectées par le VIH l'empêcher de se multiplier dans le corps, mais deux décennies de lutte contre l'épidémie ont aussi permis de faire tout un ensemble de progrès dans le sens de la résolution du problème [3].

Les inhibiteurs de la Transcriptase Inverse TI (ou reverse transcriptase) sont des molécules actives sur le VIH qui bloquent l'enzyme permettant la synthèse d'ADN (Acide DésoxyriboNucléique) complémentaire à partir de l'ARN viral (Acide RiboNucléique), avant son intégration dans le génome de la cellule infectée. Ils agissent donc sur la phase précoce de la réplication virale. Ils regroupent des molécules appartenant à des familles pharmacologiques différentes. L'approche de l'inhibition de l'enzyme transcriptase inverse (TI) du VIH s'est révélée efficace [4]. Cette enzyme traduit l'ARN viral en un ADN au sein de la cellule hôte.

Les recherches ont montré qu'un inhibiteur non nucléosidique de la transcriptase inverse appelé 1- [(2-hydroxyéthoxy) méthyl] -6- (phénylthio) thymine (HEPT) offre une grande puissance dans l'inhibition de la TI du VIH.

La chimie computationnelle désigne l'application de compétences chimiques, mathématiques et informatiques à la recherche de la solution de problèmes chimiques intéressants en utilisant des ordinateurs, et de générer des informations telles que les propriétés des molécules ou des résultats d'expériences numériques [5].

À ces fins, le développement de modèles de relations quantitatives prédictives structure-activité (QSAR) joue un rôle important dans la conception de produits chimiques fins spécifiques à un objectif, y compris les produits pharmaceutiques. La relation quantitative structure-activité (QSAR) est l'un des outils les plus pratiques en chimie physique computationnelle, Il évalue quantitativement l'activité biologique à partir de descripteurs moléculaires, il est souvent utilisé pour prédire l'activité biologique de nouveaux composés [6].

L'objectif principal de ce travail est l'application de différentes méthodes de la modélisation moléculaire pour prédire l'activité biologique anti-VIH des molécules HEPT par des descripteurs quantiques et plus particulièrement les indices de Fukui.

Le présent mémoire comporte trois chapitres :

1. Le premier chapitre traite de l'optimisation de dérivée de HEPT et les méthodes de calcul les descripteurs.
2. Le deuxième chapitre est consacré à la description de la méthodologie QSAR où nous offrons un processus d'étude QSAR ainsi une description de certaines méthodes statistiques.
3. Le troisième et le dernier chapitre présente le développement des modèles statistiques et les résultats obtenus par l'étude QSAR des 1-[2-hydroxyéthoxy-méthyl]-6-(phenylthio) thymine.

Finalement, une conclusion générale, dans laquelle nous rappelons les principaux résultats obtenus, et quelques perspectives seront proposées.

Bibliographie

- [1] Gourde, M. C. (2009). Simulacres d'une mémoire de soi : archive, deuil et identité chez Sophie Calle et Catherine Mavrikakis.
- [2] Brosset, E. (2012). Brèves observations sur un secret de Polichinelle : l'influence du droit européen sur le droit médical à travers l'exemple du secret médical.
- [3] Lévy, J. J., & Nouss, A. (1994). Sida-fiction : essai d'anthropologie romanesque. Presses Universitaires Lyon .
- [4] Boland, S. (2004). Analyse et modélisation de nouveaux inhibiteurs non nucléosidiques de la transcriptase inverse du virus de l'immunodéficience humaine de type 1 (VIH-1) (Doctoral dissertation, FUND) .
- [5] Telesnitsky A, Goff SP (1997) Reverse transcriptase and the generation of retroviral DNA. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, Retroviruses
- [6] Chtita, Samir. Modélisation de molécules organiques heterocycliques biologiquement actives par des méthodes QSAR/QSPR. Université Moulay Ismail. Meknès : s.n., 2017. Thèse.

Chapitre 1 : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH et méthodes de calculs.

Chapitre 1 : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH et méthodes de calculs

I.1 Introduction

Le VIH ou virus de l'immunodéficience humaine, est un type de virus mortelle qui peut causer une maladie appelée SIDA (syndrome d'immunodéficience acquise).

L'infection au HIV atteint le système immunitaire, qui détruit les cellules de la défense immunitaire qui provoque une baisse des défenses immunitaires.

En 1981 les premiers cas de SIDA ont été décrits aux Etats-Unis d'Amérique [1]. Selon les statistiques mondiales de 2020 il y a environ 37.6 millions de personnes vivaient avec le VIH et 690 mille de personnes sont décédées de maladies liées au SIDA [2].

Le HIV se transmet par liquide biologique qui contient le virus ou des cellules infectées par le virus, la transmission a lieu principalement par le sang, le sperme les sécrétions vaginales, et le lait maternel [3].

Jusqu'à maintenant, il y aucun traitement ne permet d'éliminer complètement le VIH de l'organisme. Les traitements adaptés permettent aux personnes infectées de bloquer la multiplication du VIH par certains inhibiteurs dans leur organisme et ainsi de garder un système immunitaire opérationnel [4].

I.2 Les inhibiteurs de la transcriptase inverse

Les inhibiteurs de la transcriptase inverse sont des molécules actives sur le VIH : ils bloquent l'enzyme permettant la synthèse d'ADN complémentaire à partir de l'ARN viral, leur principe est d'être transformés en forme active dans l'organisme et de s'intégrer à la chaîne d'ADN en formation. Les inhibiteurs de la transcriptase inverse regroupent des molécules appartenant à des familles pharmacologiques différents donc ils interrompent la chaîne de fabrication du virus [5]. Parmi les inhibiteurs connus de VIH il y a 1 - [(2-hydroxyéthoxy) méthyl] -6- (phénylthio) thymine (HEPT).

I.3 Aspect structural des dérives de HEPT

Pour la molécule HEPT, nous avons adopté la numérotation donnée dans la figure suivante, pour caractériser ses différents atomes.

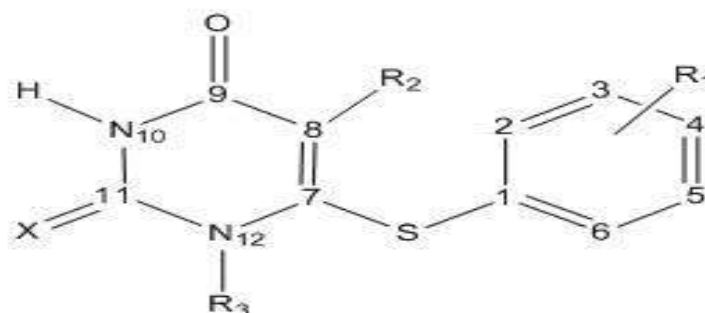


Figure I.1 Numérotation des différents atomes

Tableau I.1 L'ensemble des radicaux dans les différents dérivés de HEPT

N°	R ₁	R ₂	R ₃	X
1	2-Me	Me	CH ₂ OCH ₂ CH ₂ OH	O
2	2-NO ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O
3	2-OMe	Me	CH ₂ OCH ₂ CH ₂ OH	O
4	3-Me	Me	CH ₂ OCH ₂ CH ₂ OH	O
5	3-Et	Me	CH ₂ OCH ₂ CH ₂ OH	O
6	3-t-Bu	Me	CH ₂ OCH ₂ CH ₂ OH	O
7	3-CF ₃	Me	CH ₂ OCH ₂ CH ₂ OH	O
8	3-F	Me	CH ₂ OCH ₂ CH ₂ OH	O
9	3-Cl	Me	CH ₂ OCH ₂ CH ₂ OH	O
10	3-Br	Me	CH ₂ OCH ₂ CH ₂ OH	O
11	3-I	Me	CH ₂ OCH ₂ CH ₂ OH	O
12	3-NO ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O
13	3-OH	Me	CH ₂ OCH ₂ CH ₂ OH	O
14	3-OMe	Me	CH ₂ OCH ₂ CH ₂ OH	O
15	3,5-Me ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O
16	3,5-Cl ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O
17	3,5-Me ₂	Me	CH ₂ OCH ₂ CH ₂ OH	S
18	3-COOMe	Me	CH ₂ OCH ₂ CH ₂ OH	O
19	3-COMe	Me	CH ₂ OCH ₂ CH ₂ OH	O
20	3-CN	Me	CH ₂ OCH ₂ CH ₂ OH	O
21	H	CH ₂ CH=CH ₂	CH ₂ OCH ₂ CH ₂ OH	O
22	H	Et	CH ₂ OCH ₂ CH ₂ OH	S
23	H	Pr	CH ₂ OCH ₂ CH ₂ OH	S
24	H	<i>i</i> -Pr	CH ₂ OCH ₂ CH ₂ OH	S
25	3,5-Me ₂	Et	CH ₂ OCH ₂ CH ₂ OH	S
26	3,5-Me ₂	<i>i</i> -Pr	CH ₂ OCH ₂ CH ₂ OH	S
27	3,5-Cl ₂	Et	CH ₂ OCH ₂ CH ₂ OH	S
28	H	Et	CH ₂ OCH ₂ CH ₂ OH	O
29	H	Pr	CH ₂ OCH ₂ CH ₂ OH	O

Chapitre 1 : Dérivés HEPT pour l'inhibition de la transcriptase inverse du VIH et méthodes de calculs

30	H	<i>i</i> -Pr	CH ₂ OCH ₂ CH ₂ OH	O
31	3,5-Me ₂	Et	CH ₂ OCH ₂ CH ₂ OH	O
32	3,5-Me ₂	<i>i</i> -Pr	CH ₂ OCH ₂ CH ₂ OH	O
33	3,5-Cl ₂	Et	CH ₂ OCH ₂ CH ₂ OH	O
34	4-Me	Me	CH ₂ OCH ₂ CH ₂ OH	O
35	H	Me	CH ₂ OCH ₂ CH ₂ OH	O
36	H	Me	CH ₂ OCH ₂ CH ₂ OH	S
37	H	I	CH ₂ OCH ₂ CH ₂ OH	O
38	H	CH=CH ₂	CH ₂ OCH ₂ CH ₂ OH	O
39	H	CH=CHPh	CH ₂ OCH ₂ CH ₂ OH	O
40	H	CH ₂ Ph	CH ₂ OCH ₂ CH ₂ OH	O
41	H	CH=CPh ₂	CH ₂ OCH ₂ CH ₂ OH	O
42	H	Me	CH ₂ OCH ₂ CH ₂ OMe	O
43	H	Me	CH ₂ OCH ₂ CH ₂ OAc	O
44	H	Me	CH ₂ OCH ₂ CH ₂ OCOPh	O
45	H	Me	CH ₂ OCH ₂ Me	O
46	H	Me	CH ₂ OCH ₂ CH ₂ Cl	O
47	H	Me	CH ₂ OCH ₂ CH ₂ N ₃	O
48	H	Me	CH ₂ OCH ₂ CH ₂ F	O
49	H	Me	CH ₂ OCH ₂ CH ₂ Me	O
50	H	Me	CH ₂ OCH ₂ Ph	O
51	H	Et	CH ₂ OCH ₂ Me	O
52	H	Et	CH ₂ OCH ₂ Me	S
53	3,5-Me ₂	Et	CH ₂ OCH ₂ Me	O
54	3,5-Me ₂	Et	CH ₂ OCH ₂ Me	S
55	H	Et	CH ₂ OCH ₂ Ph	O
56	3,5-Me ₂	Et	CH ₂ OCH ₂ Ph	O
57	H	Et	CH ₂ OCH ₂ Ph	S
58	3,5-Me ₂	Et	CH ₂ OCH ₂ Ph	S
59	H	<i>i</i> -Pr	CH ₂ OCH ₂ Me	O
60	H	<i>i</i> -Pr	CH ₂ OCH ₂ Ph	O
61	H	<i>i</i> -Pr	CH ₂ OCH ₂ Me	S
62	H	<i>i</i> -Pr	CH ₂ OCH ₂ Ph	S
63	H	Me	CH ₂ OMe	O
64	H	Me	CH ₂ OBu	O
65	H	Me	Et	O
66	H	Me	Bu	O
67	3,5-Cl ₂	Et	CH ₂ OCH ₂ Me	S
68	H	Et	CH ₂ O- <i>i</i> -Pr	S
69	H	Et	CH ₂ O- <i>c</i> -Hex	S
70	H	Et	CH ₂ OCH ₂ - <i>c</i> -Hex	S
71	H	Et	CH ₂ OCH ₂ C ₆ H ₄ (4-Me)	S
72	H	Et	CH ₂ OCH ₂ C ₆ H ₄ (4-Cl)	S
73	H	Et	CH ₂ OCH ₂ CH ₂ Ph	S
74	3,5-Cl ₂	Et	CH ₂ OCH ₂ Me	O
75	H	Et	CH ₂ O- <i>i</i> -Pr	O
76	H	Et	CH ₂ O- <i>c</i> -Hex	O

77	H	Et	CH ₂ OCH ₂ -c-Hex	O
78	H	Et	CH ₂ OCH ₂ CH ₂ Ph	O
79	H	c-Pr	CH ₂ OCH ₂ Me	S
80	H	c-Pr	CH ₂ OCH ₂ Me	O
81	H	Me	CH ₂ OCH ₂ CH ₂ OC ₅ H ₁₁ - <i>n</i>	O
82	2-Cl	Me	CH ₂ OCH ₂ CH ₂ OH	O
83	3-CH ₂ OH	Me	CH ₂ OCH ₂ CH ₂ OH	O
84	4-F	Me	CH ₂ OCH ₂ CH ₂ OH	O
85	4-Cl	Me	CH ₂ OCH ₂ CH ₂ OH	O
86	4-NO ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O
87	4-CN	Me	CH ₂ OCH ₂ CH ₂ OH	O
88	4-OH	Me	CH ₂ OCH ₂ CH ₂ OH	O
89	4-OMe	Me	CH ₂ OCH ₂ CH ₂ OH	O
90	4-COMe	Me	CH ₂ OCH ₂ CH ₂ OH	O
91	3-COOH	Me	CH ₂ OCH ₂ CH ₂ OH	O
92	3-CONH ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O
93	H	COOMe	CH ₂ OCH ₂ CH ₂ OH	O
94	H	CONHPh	CH ₂ OCH ₂ CH ₂ OH	O
95	H	SPh	CH ₂ OCH ₂ CH ₂ OH	O
96	H	C≡CH	CH ₂ OCH ₂ CH ₂ OH	O
97	H	C≡C-Ph	CH ₂ OCH ₂ CH ₂ OH	O
98	3-NH ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O
99	H	COCHMe ₂	CH ₂ OCH ₂ CH ₂ OH	O
100	H	COPh	CH ₂ OCH ₂ CH ₂ OH	O
101	H	C≡CMe	CH ₂ OCH ₂ CH ₂ OH	O
102	H	F	CH ₂ OCH ₂ CH ₂ OH	O
103	H	Cl	CH ₂ OCH ₂ CH ₂ OH	O
104	H	Br	CH ₂ OCH ₂ CH ₂ OH	O
105	H	Me	CH ₂ OCH ₂ CH ₂ OCH ₂ Ph	O
106	H	Me	H	O
107	H	Me	Me	O

La Molécule principale

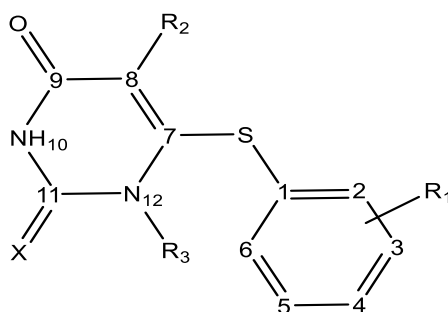


Figure I.2 1-[(2-hydroxyéthoxy) méthyl] -6- (phénylthio) thymine

Modification en R₁

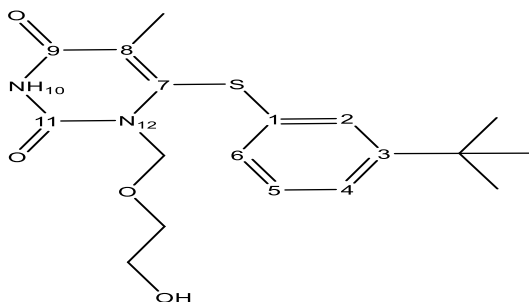


Figure I.3 HEPT 06

Modification en R₂

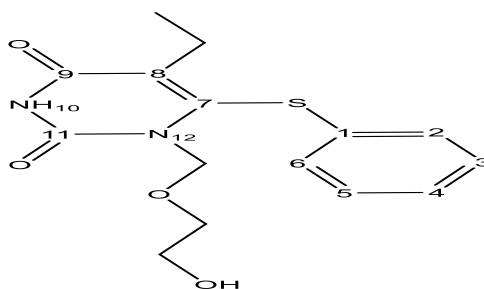


Figure I.4 HEPT 28

Modification en R₃

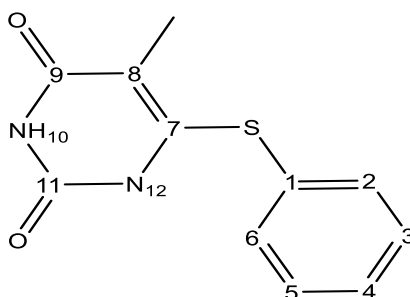


Figure I.5 HEPT 106

Modification en X

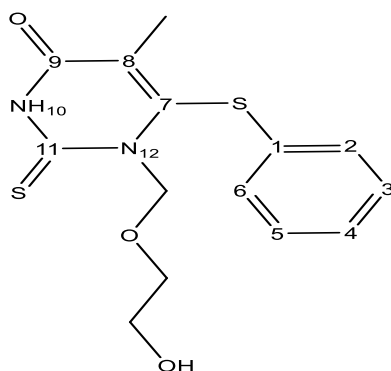


Figure I.6 HEPT 36

I.4 Activité anti-VIH des dérivés de HEPT

L'inhibition de la transcriptase inverse du VIH est l'un des différents traitements contre la maladie causée par ce virus, son rôle est d'empêcher le développement de cette dans l'organisme et de maintenir l'équilibre de l'organisme et son système de défense. On s'est basé sur l'étude bibliographique de Touhami . Il a permis d'identifier un certain nombre de dérivés de la 1-[(2 hydroxyéthoxy)méthyl]-6 (phénylthio)thymine (HEPT), qui présentent une activité inhibitrice efficace de la transcriptase inverse du VIH et sélective. L'analyse des données expérimentales ainsi recueillies montre que l'activité biologique dépend fortement des modifications de la nature, du nombre et de la position des groupements portés par le cycle thiophène et thymine. Les caractéristiques structurales et les caractéristiques électroniques sont évaluées et analysées, à savoir, le potentiel chimique, le potentiel d'ionisation, l'affinité électronique, la mollesse, la dureté, l'électronégativité, l'indice électrophile.

I.5 Les approches théoriques

I.5.1 Equation de Schrödinger

L'équation de Schrödinger indépendante du temps constitue un concept fondamental de la chimie quantique. La résolution de cette équation donne accès à l'énergie totale E ainsi qu'à la fonction d'onde Ψ associée à tout système moléculaire.

$$H\Psi = E\Psi$$

Ici, H représente l'Hamiltonien d'un système à plusieurs noyaux et électrons, il s'exprime sous la forme d'une somme d'opérateurs associés aux divers termes d'énergie cinétique et potentielle (électrostatique) :

$$H = T_e + T_n + V_{ne} + V_{ee} + V_{nn}$$

Où, en unités atomiques :

$T_e = -\sum_{i=1}^n \frac{1}{2m_i} \nabla_i^2$: Opérateur associé à l'énergie cinétique des n électrons

$T_N = -\sum_{A=1}^N \frac{1}{2M_A} \nabla_A^2$: Opérateur associé à l'énergie cinétique des N noyaux

$V_{ee} = \sum_{i=1}^n \sum_{i < j} \frac{1}{r_{ij}}$: Opérateur associé à l'énergie de répulsion électrons- électrons

$V_{NN} = \sum_{A=1}^N \sum_{A < B} \frac{Z_A Z_B}{R_{AB}}$: Opérateur associé à l'énergie de répulsion noyaux- noyaux

$V_{eN} = -\sum_{i=1}^n \sum_{A=1}^N \frac{Z_A}{r_{iA}}$: Opérateur associé à l'énergie d'attraction noyaux- électrons

Z_A désigne la charge nucléaire du noyau A, r_{iA} est la distance entre l'électron i et le noyau A.

∇_i^2 Représente l'opérateur Laplacien

L'équation (1) est la clef de voûte de la chimie quantique. Malheureusement, elle n'est soluble exactement que pour des systèmes atomiques ou moléculaires ne comprenant qu'un seul électron. Pour des systèmes poly-électroniques, on doit se contenter d'une solution approchée en introduisant des approximations portant soit sur l'Hamiltonien soit sur la forme de la fonction d'onde elle-même. L'objectif de la chimie quantique est d'obtenir de l'équation de Schrödinger une solution qui soit la plus proche possible de la solution du système physique réel [6].

I.5.2 Approximation de Born-Oppenheimer (BO)

C'est la première et la plus fondamentale des approximations utilisées dans toutes les méthodes de mécanique quantique (MQ). Elle consiste à considérer que les électrons évoluent dans un champ de noyaux fixes, dont les coordonnées $\{R_j\}$ sont des paramètres et non plus des variables du système [7]. Cette approximation se justifie par la grande différence de masse entre noyaux et électrons (un noyau est en effet 1836 fois plus lourd qu'un électron), qui fait tendre vers zéro le terme d'énergie cinétique qui leur est associé, et rend l'énergie potentielle répulsive (noyaux-noyaux) comme constante pour une géométrie donnée [8].

Dans le cadre de cette approximation, l'Hamiltonien du système est réécrit comme la somme d'un opérateur électronique et d'une constante correspondant directement à l'énergie des noyaux

$$H = H_{eN} + E_{nucl}$$

Avec :

$$H_{ele} = \sum_{i=1}^n \frac{1}{2m_i} \nabla_i^2 - \sum_{A=1}^N \sum_{i=1}^n \frac{Z_A}{r_{iA}} + \sum_{i=1}^n \sum_{i>j}^n \frac{1}{r_{ij}}$$

L'approximation de Born-Oppenheimer a ainsi permis de simplifier le problème en le réduisant à la résolution de l'équation de Schrödinger électronique suivante :

$$H_{ele} \Psi_{ele} = E_{ele} \Psi_{ele}$$

Où E_{ele} est l'énergie électronique, valeur propre de l'hamiltonien électronique H_{ele} associé à la fonction d'onde électronique Ψ_{ele} . L'énergie totale du système est calculée en fonction de l'ensemble des coordonnées des noyaux et est appelée surface d'énergie potentielle.

I.5.3 Approximation Spin-orbitales

L'évaluation de l'énergie associée aux répulsions électrons-électrons dans un système polyélectronique, demeure la principale problématique de la chimie quantique. Pour surmonter cette difficulté, il a été proposé dans un premier temps d'écrire la fonction d'onde multiélectronique comme un produit de n orbitales spatiales à un électron (Φ) (produit Hartree) [9].

$$\Psi_{ele} = \phi_1(r_1) \cdot \phi_2(r_2) \cdot \dots \cdot \phi_n(r_n) \cdot \dots$$

Néanmoins, cette description des électrons ne serait pas complète qu'avec l'ajout d'une composante de spin. Cette quantité peut avoir deux valeurs : spin $(+\frac{1}{2})$ ou spin $(-\frac{1}{2})$ auxquelles on associe respectivement les fonctions de spin $\alpha(s)$ et $\beta(s)$.

La fonction d'onde dans l'équation précédente doit tenir compte de cette coordonnée de spin. Elle est donc réécrite comme produit d'une fonction d'onde spatiale $\chi(r)$ et d'une fonction d'onde de spin $\sigma(s)$:

$$\phi(r,s) = X(r)\alpha(s) \cdot \dots \quad \text{Ou} \quad \phi(r,s) = X(r)\beta(s) \cdot \dots$$

$\Phi(r,s)$: appelée fonction spin-orbitale.

I.5.4 Approximation Hartree-Fock

Le principe d'indiscernabilité des électrons ainsi que l'antisymétrie de la fonction d'onde vis-à-vis de la permutation de deux particules ne sont pas respectés dans l'expression du produit Hartree. En effet, cette propriété d'indiscernabilité est mise au point en remplaçant les fonctions d'onde de Hartree par un déterminant de Slater [10] :

$$\Psi_{\text{Determinan}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) \phi_1(2) \dots \phi_1(N) \\ \phi_2(1) \phi_2(2) \dots \phi_2(N) \\ \dots \\ \phi_N(1) \phi_N(2) \dots \phi_N(N) \end{vmatrix}$$

Le facteur de normalisation pour la fonction est $\frac{1}{\sqrt{n!}}$ à condition que les spin-orbitales soient orthonormées, avec n le nombre d'électrons. Le déterminant de Slater satisfait bien le principe d'antisymétrie, puisqu'il change de signe lorsque l'on permute deux lignes ou deux colonnes. La meilleure fonction d'onde multiélectronique construite sur la base d'un déterminant de Slater est recherchée en minimisant l'énergie électronique à l'aide de la méthode variationnelle. Les méthodes basées sur la fonction d'onde Hartree-Fock sont désormais incapables d'évaluer l'énergie de corrélation électronique du fait que les électrons sont supposés des particules indépendantes [11,12]. Afin de remédier à cet obstacle, plusieurs méthodes ont été envisagées telles que les méthodes post-Hartree-Fock et la théorie de la fonctionnelle de la densité DFT. Dans le cadre de ce manuscrit, seule la théorie DFT basée sur la densité électronique comme entité de base sera détaillée, puisqu'il est à l'origine de tous les résultats présentés au cours de cette étude

I.5.5 La théorie de la fonctionnelle de la densité (DFT)

La Théorie de la Fonctionnelle de la Densité DFT « Density Functional Theory », est une méthode de la chimie quantique. Le point clef de cette méthode étant l'utilisation judicieuse de la fonctionnelle de la densité électronique $\rho(r)$ du système comme variable fondamentale définissant l'énergie. La densité électronique sur laquelle repose cette théorie est seulement fonction de 3 variables (x, y, z); ce qui réduit considérablement les temps de calculs et permet d'étudier des systèmes de grandes tailles hors de portée des autres méthodes ab-initio basées sur la fonction d'onde à $3N$ variables (où N est le nombre de particules du système). Le principe de cette approche est d'évaluer l'énergie de l'état fondamental d'un système à n électrons soumis à un potentiel externe $v(r)$ via une formulation variationnelle impliquant la densité électronique

$\rho(r)$. En se basant sur la notion de la densité électronique, plusieurs tentatives de mise en place d'un formalisme quantique ont été proposées [13,14]. Les premiers concepts de la DFT ont été validés en 1964 par Hohenberg et Kohn [15], qui ont démontré qu'à une densité électronique donnée est associée un unique Hamiltonien et réciproquement. Ensuite, Kohn et Sham [16] ont fait intervenir une fonctionnelle dite d'échange-corrélation qui modélise la corrélation électronique dans les calculs de chimie quantique. Le but des méthodes DFT est de trouver une fonctionnelle (Une fonctionnelle associe un scalaire à une fonction : c'est une fonction de fonction) permettant de connecter la densité avec l'énergie.

I.5.5.1 Fondement de la théorie DFT

Plusieurs approximations sont à l'origine du succès de la DFT. Elles traitent toutes de la fonctionnelle d'échange-corrélation ; elles sont généralement notées, en pratique, par un groupe de lettres désignant les initiales des auteurs dont la première partie est relative à la méthode de calcul du terme d'échange et la deuxième, celle du terme de corrélation [17].

I.5.5.2 Approximation locale de la densité (Local Density Approximation, LDA)

L'approximation locale de la densité (LDA), où le fonctionnel dépend seulement de la densité locale fut proposée par Kohn et Sham. L'idée de base est qu'il est possible d'estimer l'énergie d'échange-corrélation d'un système inhomogène en utilisant, sur des portions infinitésimales, les résultats d'un gaz homogène d'électrons de densité égale à la densité locale du système réel [18]. La somme de toutes les contributions ponctuelles permet d'obtenir l'énergie d'échange corrélation selon l'équation suivante :

$$E_{xc}[p] = \int P(r) \varepsilon_{xc}[P(r)] d^3r$$

Où ε_{xc} représente l'énergie d'échange corrélation d'un gaz d'électron uniforme. Pour les systèmes magnétiques, la LDA doit être étendue à l'Approximation de la Densité Locale de Spin (LSDA : Local Spin Density Approximation), où l'énergie d'échange et corrélation est une fonctionnelle des deux densités de spin haut et bas :

$$[p_{\uparrow}, p_{\downarrow}] = \int P(r) \varepsilon_{xc}[p_{\uparrow}(r), p_{\downarrow}(r)] d^3r$$

La LDA suppose que cette fonctionnelle est purement locale. Cette énergie est divisée en deux termes :

$$\varepsilon_{xc} = \varepsilon_c(p) + \varepsilon_x(p)$$

ε_x Est l'énergie d'échange et ε_c est l'énergie de corrélation.

La fonctionnelle xc e peut être constante, mais généralement, elle est déterminée par des procédures de paramétrage comme celles de Wigner [19], Ceperly et Alder [20], Perdew et Zunger [21], Hedin et Lundqvist [22], et Perdew et Wang [23].

Pour les systèmes magnétiques, la LDA doit être étendue à l'approximation de la densité locale de Spin (LSDA Local Spin Density Approximation), où l'énergie d'échange et corrélation est une fonctionnelle des deux densités de spin haut et bas.

I.5.5.3 Approximation du Gradient Généralisé(GGA)

Cette méthode consiste à introduire le gradient de la densité dans la description des effets d'échange et de corrélation. Alors que la méthode LDA est locale de par la considération d'une densité équivalente à celle d'un gaz homogène, on incorpore ici des éléments non-locaux. En effet, le gradient de la densité permet de prendre en compte la variation de la densité au voisinage de chaque point.

L'énergie d'échange-corrélation, dans le cadre de l'approximation du gradient généralise GGA (Generalized-gradient approximation) (l'anglais Generalized Gradient Approximation) [24], s'écrit alors :

$$E_{XC}^{GGA} = [\rho_{\alpha}(r), \rho_{\beta}(r)] = \int \varepsilon_{xc}^{GGA}(\rho(r), \nabla\rho(r)) dr$$

Où ε_{xc}^{GGA} :est la densité d'énergie d'échange-corrélation.

E_{XC}^{GGA} : Est divisé en deux contributions échange et corrélation.

$$E_{XC}^{GGA} = E_X^{GGA} = E_C^{GGA}$$

I.5.5.4 Fonctionnelle hybride B3LYP

Une combinaison entre la fonctionnelle d'échange B88 proposée par Becke en 1988 [25] et la fonctionnelle de corrélation LYP développée par Lee, Yang et Parr [26] a donné naissance à la fonctionnelle DFT : BLYP

Quelques années plus tard, en s'appuyant sur le principe de la fonctionnelle hybride, Stephens et coll [27] ont proposé la fonctionnelle B3LYP qui inclut 20% d'échange Hartree-Fock dans la fonctionnelle d'échange.

L'expression de la fonctionnelle hybride B3LYP intégrée dans le logiciel Gaussian 09 est de la forme suivante :

$$E_{B3LYP}^{xc} = (a)E_{HF}^x + (1+a)E_{LSDA}^x + (b)E_{GGA}^c + E_{GGA}^c + (1-c)E_{LSDA}^c$$

Les trois paramètres (a, b et c) sont des coefficients semi empirique (d'où le numéro 3 dans B3LYP) optimisés par Becke pour adhérer à des données expérimentales de systèmes standards servant de tests. De l'équation (14 I), il s'avère que B3LYP est une fonction hybride avec 20% d'échange Hartree-Fock (aE_{HF}^x)

B3LYP est la fonctionnelle hybride la plus connue et abondamment la plus utilisée des fonctionnelles citées en littérature, si bien qu'aujourd'hui, le terme DFT est presque synonyme de réaliser un calcul B3LYP. Néanmoins, cette méthode présente quelques limitations telles que:

- La sous-estimation des hauteurs de barrière énergétique.
- La sous-estimation des énergies correspondantes aux interactions à long terme telles que les interactions van der Waals.

I.6 Calcul des descripteurs quantiques

I.6.1- Les indices de Fukui

La fonction de Fukui, définie par l'équation

$$\left(\frac{\delta^2 E}{\partial N \delta v(\vec{r})} \right) = \left(\frac{\partial \rho(\vec{r})}{\partial N} \right)_v = \left(\frac{\delta \mu}{\delta v(\vec{r})} \right)_N = f(\vec{r})$$

C'est un descripteur de réactivité central de la DFT. Elle traduit la propension de la densité électronique à se déformer en r sous l'effet d'une modification du nombre d'électrons du système. Elle renseigne sur la façon la plus énergétiquement favorable de modifier le nombre d'électrons du système.

Il est courant de distinguer le cas d'un gain d'électron et celui d'une perte d'électron, Ainsi f^+ est la réponse de la densité électronique à un gain d'électron (ou à une attaque nucléophile) à potentiel externe constant tandis que f^- est la réponse à une perte d'électron (ou à une attaque électrophile) à potentiel externe constant. Il est important de remarquer que compte-tenu de leurs définitions, f^+ et f^- sont des grandeurs positives. Ainsi au cours d'un transfert d'électron d'un nucléophile vers un électrophile, ce sont les sites de l'électrophile (E) où f^+ est la plus grande qui vont voir leur densité électronique augmentée considérablement et ce sont les sites du nucléophile (Nu) où f^- est la plus grande qui vont perdre le plus de densité électronique.

$$f^+(r) = \left(\frac{\partial \left(\frac{\delta E}{\delta v(r)} \right)_N}{\partial N} \right)_{v(r)} = \left(\frac{\delta \left(\frac{\partial E}{\partial N} \right)_{v(r)}^+}{\delta v(r)} \right)_N$$

$$f^+(r) = \rho_{N+1,v(r)}(r) - \rho_{N,v(r)}(r)$$

$$f^-(r) = \left(\frac{\partial \left(\frac{\delta E}{\delta v(r)} \right)_N}{\partial N} \right)_{v(r)}^- = \left(\frac{\delta \left(\frac{\partial E}{\partial N} \right)_{v(r)}^-}{\delta v(r)} \right)_N$$

$$f^-(r) = \rho_{N,v(r)}(r) - \rho_{N-1,v(r)}(r)$$

I.6.2 Potentiel d'ionisation (D1)

C'est l'énergie nécessaire pour arracher un électron d'un système. C'est-à-dire l'énergie nécessaire pour passer de la molécule neutre (N électrons) au cation (N- 1 électrons) :

$$I = E(N-1) - E(N)$$

I.6.3 Affinité électronique (D2)

C'est l'énergie gagnée par un système lorsqu'il capte un électron. C'est à dire le gain d'énergie qu'accompagne le passage d'un système neutre à un anion :

$$A = E(N) - E(N+1)$$

I.6.4 La dureté chimique et la mollesse globale (D3)

La première dérivée partielle de μ par rapport à N (le nombre total d'électrons) est définie comme la dureté (*hardness*) globale η du système :

$$2\eta = \left[\frac{\partial \mu}{\partial N} \right]_{v(r)} = \left[\frac{\partial^2 E}{\partial N^2} \right]_{v(r)} = \frac{1}{S}$$

Avec :

➤ S : la mollesse (*softness*) globale du système

η et S peuvent être écrits comme suit :

$$\eta = \frac{I - A}{2}$$

$$S = \frac{1}{I - A}$$

I.6.5 Indice d'électrophilicité global (D4)

L'indice d'électrophilicité ω est lié au potentiel chimique μ par la relation suivante :

$$\omega = \frac{\mu^2}{2\eta}$$

Cet indice exprime la capacité d'un électrophile d'acquérir une charge électronique supplémentaire [28].

I.7 Conclusion

Dans ce chapitre on a réalisé des calculs d'optimisation de géométrie des dérivés HEPT au niveau DFT avec le fonctionnelle B3LYP et la base 6-311G+(d,p). A partir des résultats nous avons calculé les descripteurs suivants : le potentiel d'ionisation, l'affinité électronique, la mollesse, l'indice électrophile et les indices de Fukui.

Bibliographie

- [1] Burack, JH, Barrett, DC, Stall, RD, Chesney, MA, Ekstrand, ML et Coates, TJ (1993). Symptômes dépressifs et déclin des lymphocytes CD4 chez les hommes infectés par le VIH. *Jama*, 270 (21), 2568-2573.
- [2] Edward R. Cachay, MD, MAS, University of California, San Diego Dernière révision totale août 2019/ Dernière modification du contenu août 2019
- [3] [http://inv.santepubliquefrance.fr/Centre medicale de l'institut pasteur](http://inv.santepubliquefrance.fr/Centre_medicale_de_l_institut_pasteur) (Paul-Henri Conigny) décembre 2018
- [4] étude de l'efficacité et la tolérance de la TRIOMUNE : Bilan de douze mois de suivi dans le service des maladies infectieuses et tropicales du CHU du point G Mr Souleymane Dramane COULIBALY
- [5] traitement contre le virus de SIDA professeur ANNE-MARIE LEHER DRYLEWICZ médecin le 2010
- [6] Gilbert, A. (2007). Introduction à la chimie quantique computationnelle : théorie. Curso : ANU-The Australian National University .
- [7] Chabbal, S. (2011). Formalisme et implémentation des gradients analytiques pour les méthodes mixtes sr-DFT/lr-WFT et leurs applications à quelques cas critiques de la DFT.
- [8] Desrier, A. (2018). Dynamique ultrarapide corrélée: théorie, simulations et interprétations d'expériences de spectroscopie " attoseconde" (Doctoral dissertation, Sorbonne université).
- [9] Bouchelouh, S., & Hala, A. (2020). Etude théorique des propriétés physico-chimiques des flavonoïdes d'une plante (salvia officinalis) (Doctoral dissertation, University of Jijel).
- [10] McDougall, R. (2001). Numérisation des équations de Hartree-Fock (Doctoral dissertation, Université du Québec à Trois-Rivières).
- [11] Annexe, B. (2017). Hartree-Fock. École doctorale des sciences chimiques-ED250 Institut de Chimie Radicalaire-UMR 7273, 199.
- [12] Fock, V. (1930). Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik*, 61(1-2), 126-148.
- [13] Thomas, LH (1927, janvier). Le calcul des champs atomiques. Dans les procédures mathématiques de la société philosophique de Cambridge (Vol. 23, No. 5, pp. 542-548). La presse de l'Universite de Cambridge.
- [14] Fermi, E. (1927). Méthode statistique pour déterminer certaines propriétés des atomes. *Déchirer. Accad. Naz. Lincei*, 6 (602-607), 5.

- [15] Kepple, P., & Griem, HR (1968). Calculs de profil Stark améliorés pour les raies d'hydrogène H , H β , H γ et H δ . *Physical Review* , 173 (1), 317.
- [16] Kohn, W. (1996). Fonction de densité et méthode de matrice de densité mise à l'échelle linéairement avec le nombre d'atomes. *Physical Review Letters* , 76 (17), 3168.
- [17] Neto, R. P. (2014). Développement et application de méthodes corrélées pour la description de systèmes moléculaires (Doctoral dissertation, Université Pierre et Marie Curie-Paris VI).
- [18] Payne, M. C. (1992). „MP Teter, DC Allan, TA Arias, JD Joannopoulos. *Rev. Mod. Phys.*, 64, 1045.
- [19] Barrera, RG, Grether, M., & de Llano, M. (1979). États HF d'ordre à longue distance dans le modèle de gelée déformable. *Journal of Physics C: Solid State Physics* , 12 (18), L715.
- [20] Ceperley, DM, & Alder, BJ (1980). Etat fondamental du gaz d'électrons par une méthode stochastique. *Lettres d'examen physique* , 45 (7), 566.
- [21] Perdew, JP et Wang, Y. (1992). Représentation analytique précise et simple de l'énergie de corrélation électron-gaz. *Physical Review B* , 45 (23), 13244.
- [22] Hedin, L., & Lundqvist, BI (1971). Potentiels locaux explicites de corrélation d'échange. *Journal of Physics C: Physique du solide* , 4 (14), 2064.
- [23] Perdew, JP et Wang, Y. (2018). Erratum : Représentation analytique précise et simple de l'énergie de corrélation électron-gaz [Phys. Rév. B 45, 13244 (1992)]. *Physical Review B* , 98 (7), 079904.
- [24] Schmider, HL et Becke, AD (1998). Fonctionnalités de densité optimisées de l'ensemble de test G2 étendu. *Le Journal de physique chimique* , 108 (23), 9624-9631.
- [25] Becke, AD (1988). Approximation d'échange-énergie densité-fonctionnelle avec un comportement asymptotique correct. *Physical review A* , 38 (6), 3098.
- [26] Zhu, T., Lee, C. et Yang, W. (1993). Examen de plusieurs fonctionnelles énergétiques d'échange-corrélation par des calculs atomiques précis et auto-cohérents. *Le Journal de physique chimique* , 98 (6), 4814-4821.
- [27] Stephens, PJ, Devlin, FJ, Chabalowski, CF et Frisch, MJ (1994). Calcul ab initio des spectres d'absorption vibrationnelle et de dichroïsme circulaire à l'aide de champs de force fonctionnels de densité. *Le Journal de chimie physique* , 98 (45), 11623-11627.
- [28] Labet, V. (2009). Etude Théorique de Quelques Aspects de la Réactivité des Bases de l'ADN-Définition de nouveaux outils théoriques d'étude de la réactivité chimique (Doctoral dissertation, Université Joseph-Fourier-Grenoble I).

[29] TOUHAMI Moufida. 2019. Modélisation de l'activité biologique de composés hétérocycles. Thèse de doctorat. Université de Saida. Algérie.

**Chapitre 2 : Relations Quantitatives
Structures Activités/Propriétés
(QSAR/QSPR)**

Chapitre 2 : Relations Quantitatives Structures Activités/Propriétés (QSAR/QSPR)

II.1 Introduction

Le développement de modèles de relations quantitatives prédictives structure-activité (QSAR) joue un rôle important dans la conception de produits chimiques fins spécifiques à un objectif, y compris les produits pharmaceutiques [1].

Ces méthodes s'appuient sur le principe que les propriétés physico-chimiques et les activités biologiques des molécules dépendent fortement de leurs structures chimiques [2]. La modélisation QSAR est un outil utile pour la prédiction de l'activité biologique, des propriétés physico-chimiques et des réponses toxicologiques de composés chimiques non testés.

Ce chapitre vise à couvrir les concepts et techniques essentiels qui sont pertinents pour la réalisation, le développement et de validation d'études QSAR / QSPR.

II.2 Modélisation QSAR/QSPR

La modélisation moléculaire est un outil pour les chercheurs qui étudient la structure moléculaire et la réactivité. Connaissance de la structure du bâtiment Les molécules permettent de comprendre les résultats obtenus en transformation physique, chimique ou biologique [3]. La modélisation moléculaire implique l'utilisation de méthodes de calcul théoriques avec différentes méthodes (mécanique moléculaire, méthodes semi-empiriques ou quantiques (ab-initio, DFT, ...), permettant de déterminer la représentation graphique de la géométrie et d'évaluer les propriétés physiques et chimiques de la molécule à l'étude. La modélisation moléculaire permet d'expliquer des phénomènes physiques et chimiques, de proposer de nouvelles expériences et d'analyser les résultats de manière plus critique que les expériences traditionnellement utilisées, c'est donc un complément aux expériences [4].

Le développement de nouvelles techniques de modélisation a permis la mise en place de nombreuses méthodes QSPR/QSAR (Quantitative Structure Property/activity Relationship); elles reposent pour la plupart sur la recherche d'une relation entre un ensemble de nombres réels, appelés descripteurs moléculaires, et la propriété ou l'activité que l'on souhaite prédire.

Ces méthodes permettent d'interpréter les données expérimentales disponibles et de prédire les propriétés/activités pour des nouveaux composés ou des composés pour lesquels les données expérimentales ne sont pas disponibles [5].

II.3 Définition et formalisme

QSAR/QSPR est une corrélation mathématique entre une activité biologique ou une propriété moléculaire spécifiée et une ou plusieurs propriétés structurelles physicochimiques et/ou moléculaires, appelées descripteurs puisqu'elles "décrivent" l'activité ou la propriété examinée [6].

$$\text{Propriété / Activité} = f(\text{descripteurs moléculaires})$$

II.4 Objectifs du QSAR

Les objectifs pratiques d'une étude QSAR sont nombreux et ces techniques sont largement utilisées dans de nombreuses situations. Le but des études QSAR comprend donc ce qui suit [7]:

- Prédire l'activité biologique et les propriétés physico-chimiques par des moyens rationnels.
- Économies sur les coûts de développement de produits.
- Les prévisions pourraient réduire la nécessité de tests sur les animaux longs et coûteux.
- Ainsi que l'analyse des données structurelles afin de révéler les déterminants de la propriété ou de l'activité mesurée.

II.5 Les étapes de développement des modèles QSAR

L'étude QSAR peut être visualisée comme comprenant trois étapes simples, à savoir (a) la préparation des données, (b) le traitement des données et (c) l'évaluation des résultats.

II.5.1 La préparation des données

En QSAR/QSPR, l'ensemble de données représente chaque composé chimique. Les données quantitatives sont obtenues à partir de deux composants principaux, à savoir la réponse ou le critère d'évaluation à traiter (activité/propriété) et le prédicteur ou les variables indépendantes appelés descripteurs définissant les attributs chimiques.

Les descripteurs

En chimie un descripteur est une propriété structurale pour caractériser une molécule et qui permet de comparer les différentes molécules et la recherche de molécules apparentées dans une base de données. C'est une représentation mathématique formelle d'une molécule capable de transformer l'information chimique codée à l'intérieur d'une représentation symbolique de la même molécule, en une valeur numérique utile [8].

Les descripteurs moléculaires sont utilisés dans la caractérisation des propriétés de composés chimiques dans des domaines scientifiques tels que toxicologie, chimie analytique, chimie pharmaceutique et chimie de l'environnement.

Historiquement, deux grands schémas pour la classification des descripteurs moléculaires ont été établis : l'un en fonction de leur origine (constitutionnel, topologique, géométrique, quantique, thermodynamique...), et le deuxième en fonction de leur dimensionnalité (1D, 2D, 3D) [9].

Les type de descripteurs

- **Les descripteurs 1D** : Il peut être obtenu à partir de la formule moléculaire (par exemple, C₆H₆O du phénol) et décrit les propriétés globales du composé. Par exemple, il s'agit de sa composition, des atomes qui le composent ou de sa masse molaire. On peut noter que ces descripteurs ne permettent pas de distinguer les isomères constitutifs et ne permettent pas le développement de modèles plus complexes. Besoin d'ajouter d'autres types de descripteurs [10]. Des exemples de descripteurs constitutionnels sont
 - La masse molaire
 - Les nombres absolus et relatifs d'atomes (C, H, O, S, N, F, Cl, Br, I, P. . .).
 - Les nombres absolus et relatifs de groupes fonctionnels (NH₂, COOH, OH. . .).
 - Les nombres absolus et relatifs de liaisons (simples, doubles, aromatiques. . .).
 - Les nombres absolus et relatifs de cycles (aromatiques ou non) [11].
- **Les descripteurs 2D** : Les descripteurs 2D sont des attributs numériques qui peuvent être calculés à partir de la table de connexion des molécules ou de la représentation plane (2D) de la structure. Ils sont basés sur la présence d'éléments, les charges partielles, la nature des liaisons et le nombre de cycles. La métrique la plus couramment utilisée est le coefficient de partage octanol-eau P, qui est généralement utilisé sous sa forme logarithmique (log P) [12].

$$\log P = \log \frac{[\text{octanol}]}{[\text{H}_2\text{O}]}$$

- **Les descripteurs 3D** : Les descripteurs 3D des molécules sont évalués en fonction des positions relatives de leurs atomes dans l'espace et décrivent des caractéristiques plus complexes ; par conséquent, leurs calculs doivent comprendre la géométrie 3D de la molécule, généralement par modélisation moléculaire. Par conséquent, ces descripteurs s'avèrent relativement coûteux en temps de calcul, mais fournissent plus d'informations et sont nécessaires pour modéliser des propriétés ou des activités qui dépendent de structures 3D. Les propriétés électroniques et l'un parmi plusieurs familles importantes de descripteurs 3D [13].

Par exemple, les énergies de la plus haute orbitale moléculaire occupée HOMO et de la plus basse vacante LUMO, le moment dipolaire, la polarisabilité, le potentiel d'ionisation

Lors de la préparation des données QSAR, il faut veiller à ce que les structures moléculaires aient été correctement dessinées ou importées, que les données d'activité biologique (ou autre réponse) proviennent d'une source authentique et que le descripteur les valeurs ont été calculées à l'aide d'un logiciel validé.

II.5.2 Le traitement des données

L'étape de traitement des données comprend généralement plusieurs opérations de prétraitement avant le développement du modèle, telles que la division des données et le développement des modèles QSAR/QSPR.

La division des données

La division des données est la sélection des ensembles d'apprentissage et de test. En général, la division de l'ensemble de données en ensembles d'apprentissage et de test doit être exécutée de telle manière que les données représentant à la fois les ensembles d'apprentissage et de test soient dispersés dans tout l'espace de descripteur occupé par l'ensemble de données entier et chaque point de l'ensemble de test est proche d'au moins un composé de l'ensemble d'apprentissage. Plusieurs approches sont utilisées pour la sélection des ensembles d'apprentissage et de test tels que : sélection aléatoire, k-means,

Les études QSAR/QSPR visent à développer des modèles de corrélation utilisant une réponse des produits chimiques (activité/propriété) et des données d'information chimique (descripteurs) dans une approche statistique.

Des approches basées sur la régression sont utilisées lorsque les données de réponse des produits chimiques sont entièrement numériques, c'est-à-dire quantitatives, tandis que les réponses chimiques qualitatives ou semi-quantitatives sont modélisées à l'aide de techniques de classification.

Un grand nombre de techniques statistiques sont disponibles pour réaliser une étude QSAR, allant de l'analyse de régression, telles que analyse en composantes principales (ACP) et les Réseaux de Neurones Artificielle (RNA).

▪ L'analyse de régression

L'analyse de régression est une méthode statistique de modélisation des relations entre différentes variables (dépendantes et indépendantes). Elle est utilisée pour décrire et analyser les relations entre les données [14].

Dans la modélisation QSAR, la régression prend généralement la forme suivante :

$$\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Où

- $\beta_0, \beta_1, \beta_2$ sont appelés coefficient de régression
- β_0 est appelé interception.
- $X_1 \dots X_k$ sont des descripteurs moléculaires avec leur coefficients correspondants β_k .

L'analyse de régression a un certain nombre d'avantages dans QSAR car il est simple à réaliser, et fournit des modèles clairs, et transparente sans ambiguïté.

▪ **L'analyse en composantes principales (ACP)**

L'Analyse en Composantes Principales (ACP) (PCA, pour Principal Component Analysis), est une méthode d'analyse de données, qui cherche à synthétiser l'information contenue dans un tableau croisant des individus et des variables quantitatives [15]. C'est une technique qui est utilisée pour réduire la dimension de l'espace de représentation des données, il consiste à remplacer les variables initiales par de nouvelles variables, appelées composantes principales, deux à deux non corrélées, et telles que les projections des données sur ces composantes soient de variance maximale.

▪ **Les Réseaux de Neurones Artificielle**

Un réseau de neurones artificiels est un système dont la conception est à l'origine schématiquement inspirée du fonctionnement des neurones biologiques, et qui par la suite s'est rapproché des méthodes statistiques, Ils sont particulièrement utilisés pour résoudre des problèmes de classification, de prédiction, de reconnaissance des formes, de catégorisation, et d'optimisation C'est une technique de reconnaissance qui ressemble de près le fonctionnement interne du cerveau qui est essentiellement composé de neurones interconnectés.

Les réseaux de neurones artificiels sont constitués de nœuds ou d'unités de traitement connectés entre eux pour former un réseau de nœuds, et chaque connexion entre ces nœuds possède un ensemble de valeurs appelées poids qui contribuent à déterminer les valeurs résultant de chaque élément de traitement en fonction des valeurs entrée pour cet élément [17].

II. 5. 3 L'évaluation et la validation des résultats

Le défi dans le processus de développement d'un modèle QSAR n'est plus de développer un modèle statistiquement valable pour prédire l'activité au sein de l'ensemble d'apprentissage, mais de développer un modèle capable de prédire avec précision l'activité de nouveaux produits chimiques.

Des outils de validation doivent être appliqués sur les modèle QSAR développé pour vérifier leur prédictivité pour de nouvelles molécules non testées.

- **La validation interne :** La validation interne d'un modèle QSAR est effectuée sur la base des molécules utilisées dans le développement du modèle. Il s'agit d'une prédiction

d'activité des molécules étudiées suivie d'une estimation de paramètres permettant de détecter la précision des prédictions. L'inconvénient majeur de cette approche est le manque de prévisibilité du modèle lorsqu'il est appliqué à un nouvel ensemble de données.

- **La validation externe :** pour la validation externe, l'ensemble de données est généralement divisé en ensembles d'apprentissage et de test. Le modèle est initialement développé sur l'ensemble d'apprentissage, puis il est utilisé pour vérifier la validation externe en utilisant les molécules de l'ensemble de test qui ne sont pas utilisées.

II.6 Les métriques d'évaluation

Les métriques d'évaluation les plus couramment utilisées sont discutées ci-dessous :

- **Le coefficient de corrélation (R^2) :** reflète le degré de corrélation entre les valeurs observées et prédites les données d'activité de l'ensemble de test.

$$R^2 = 1 - \frac{\sum (Y_{\text{obs}} - Y_{\text{calc}})^2}{\sum (Y_{\text{obs}} - \overline{Y_{\text{obs}}})^2}$$

Où :

- Y_{obs} : La valeur expérimentale de l'activité
- Y_{calc} : La valeur calculée de l'activité
- $\overline{Y_{\text{obs}}}$: La valeur moyenne des valeurs calculées de l'activité.
- **R^2 ajusté :** R^2 ajusté indique dans quelle mesure les attributs s'ajustent à une courbe ou une ligne, mais s'ajuste au nombre de termes dans un modèle. Si vous ajoutez de plus en plus de variables inutiles à un modèle, le R^2 ajusté diminuera. Si vous ajoutez des variables plus utiles, le R^2 ajusté augmentera. R^2 ajusté sera toujours inférieur ou égal à R^2 .

$$R^2_{\text{ajusté}} = R^2 - \frac{k(1-R^2)}{N-K-1}$$

Où :

- N : nombre de molécules
- R^2 : la corrélation
- K : le nombre de descripteurs
- **La Précision :** La précision représente le nombre de données classées correctement

comme positives par exemple, par rapport au nombre de données totales reconnues comme positives :

▪

$$precision = \frac{VP}{VP + FP}$$

Où :

- VP : Varies positive
- FP : Faux positive

II.7 Les applications du QSAR

Les applications des méthodes QSAR sont très nombreuses, elles touchent tous les domaines où la structure chimique intervient, entre autre on peut citer [18] :

- La prédiction d'une variété de propriétés physico-chimique des molécules.
- Prédiction de la toxicité aquatique des composés chimiques vis-à-vis des espèces environnementales.
- Prédiction de plusieurs propriétés dangereuses telle que l'explosibilité et l'inflammabilité de certaines familles de molécules chimiques.
- Conception des médicaments et de nombreux autres produits tels que les agents tensio-actifs, parfums, les colorants et les produits de la chimie fine.

II.8 Conclusion

La technique de modélisation prédictive QSAR offre une option pour développer une base mathématique pour les réponses chimiques suscitées. Actuellement les méthodes QSAR/QSPR sont devenues de plus en plus utilisées dans le cadre de conception de nouvelles entités chimiques.

La modélisation QSAR/QSPR implique l'utilisation d'un nombre important d'outils statistiques. Le choix de la méthode de développement dépend d'une multitude de facteurs, y compris la réponse à modéliser, la nature des données de l'ensemble d'apprentissage, le type de descripteurs utilisés et également ses nombres, et même l'objectif de l'analyse.

Dès lors que les relations mathématiques quantitatives sont établies, la validation des modèles à l'aide d'un algorithme statistique approprié devient essentielle pour confirmer la stabilité et la prédictivité des modèles.

Dans ce chapitre nous avons présenté une brève présentation des principales étapes de développement et générations de modèles QSAR/QSPR.

Bibliographie

- [1] B. K. ERRAHOUÏ, Etude des relations quantitatives structure–toxicité des composés chimiques à l’aide des descripteurs moléculaires. « Modélisation QSAR », UNIVERSITE ABOU BEKR BELKAÏD DE TLEMCEÏN, 2015.
- [2] Mabrouk Hamadache, « A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, Domain of Application and Prediction,» Journal of Hazardous Materials, Elsevier, pp. 28-40, 2016.
- [3] G.Segal. La modélisation moléculaire-Les logiciels : tendances et évolution. Le Technoscope de Bio futur, n °34, février (1990).
- [4] S. K. Niazi, Preformulation Handbook of Chemical, Pharmaceutical Scientist Inc.,USA, 2007.
- [5] A.F.A Cros, Action de l'alcool amylique sur l'organisme, thèse de doctorat, faculté de Médecine, Université Strasbourg , Strasbourg , 1863.
- [6] K. Roy, «Advances in QSAR Modeling: Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental sciences,» pp. 57-58.
- [7] Mark T. D. Cronin, Recent Advances in QSAR Studies: Methods and Applications
- [8] Tomasz Puzyn, Jerzy Leszczynski, Mark T. D. Cronin, Les progrès récents dans QSAR études: méthodes et applications, Springer science Business Media, 2010, p. 30, ISBN 1-4020-9783-2.
- [9] Abiza-Fiala. Etude par modélisation moléculaire l’activité cytotoxiques d’une série de dérivés de benzimidazole (QSAR et Docking moléculaire) Mémoire de fin d’étude en Master 2018.
- [10] S.W. Benson, J.H. Buss, J. Chem. Phys., 1958, 29, 546-572.
- [11] Wiener H., Structural determination of paraffin boiling points. Journal of Chemical Information and Computer Sciences, 69, (1947), 17.
- [12] P. Geerlings, F. De Proft, W. Langenaeker, Chem. Rev., 2003, 103, 1793-1874.
- [13] P. Labute, A widely applicable set of descriptors, Journal of Molecular Graphics and Modelling, 18, 2000, 464–477.
- [14] H. Van de Waterbeemd, (ed.). (1995). Chemometric Methods in Molecular Design, Weinheim, Germany: VCH.
- [15] C. Chatfield and A.J. Collins, Introduction to Multivariate Analysis, Chapman and Hall, 1980.
- [16] H. Wold, Multivariate analysis, Academic Press, 1966.

[17] P.J. DREW, et J.R.T. MONSON, (2000). Artificial neural networks, *Surgery* 127: 3-11.

[18] Katritzky, «Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction,» *Chem. Rev.*, vol. 110, p. 5714–5789, 2010.

Chapitre 3 : Résultats et Discussions

Chapitre 3 : Résultats et Discussions

III.1 Introduction

L'analyse QSAR est maintenant largement utilisée comme outil rationnel pour la prédiction et la conception de produits chimiques bénéfiques pour la santé, de processus industriels/de laboratoire ou d'applications domestiques. Les études QSAR/QSPR visent à développer des modèles de corrélation entre une activité/propriété de produits chimiques et des données d'informations chimiques par des approches statistiques (descripteurs).

Dans ce travail, nous intéressons à l'étude des propriétés physico-chimiques des HEPT qui permettent de prévoir les paramètres physicochimiques qui sont influents sur l'activité biologique, afin de prédire l'activité biologique de nouvelles molécules.

Divers outils d'apprentissage automatique sont utiles pour l'analyse de modélisation QSAR/QSPR. Cependant, avant la construction des modèles QSAR ; un processus de calcul des descripteurs suivi de l'application d'une méthode de sélection de descripteurs appropriée doivent être prises pour s'assurer que toutes les expériences effectuées pour avoir un bon modèle de prédiction.

Dans ce chapitre, des modèles QSAR de l'inhibition de la transcriptase inverse de dérivés HEPT ont été développés et les résultats sont ensuite discutés.

III.2 Préparation et prétraitement des données

Un modèle QSAR peut être exprimé sous la forme d'une équation mathématique simple qui peut corréler les propriétés physicochimiques ou biologiques de molécules utilisant divers paramètres quantitatifs dérivés par calcul ou expérimentation appelés « descripteurs ». Les descripteurs sont corrélés avec les propriétés/activités expérimentales afin d'obtenir un modèle QSAR statistiquement significatif. Les descripteurs jouent un rôle crucial dans le développement de tout modèle QSAR car ils représentent quantitativement l'information chimique codée.

Dans ce travail, la plupart des molécules du dérivé 1-[(2-hydroxyéthoxy)méthyl]-6(phénylthio)thymine ont été pré-optimisées en utilisant la ligne de base 6-31G pour la fonction B3LYP qui est une sorte de méthode de théorie fonctionnelle de la densité DFT. Les structures de départ sont obtenues en utilisant logiciel GaussView 09 par la modification de la molécule principale HEPT on change les radicaux R1, R2, R3 et X, à la fin de ces calculs, des propriétés électroniques seront obtenues pour calculer la liste des descripteurs [1].

Les descripteurs que nous avons utilisés dans notre travail sont les suivants :

- **D1** : le potentiel d'ionisation (I)

Chapitre 3 : Résultats et Discussions

- **D2** : l'affinité électronique (A)
- **D3** : la Mollesse globale (S)
- **D4** : l'indice d'électrophilicité (ω)
- **D5** : descripteur dual du site C7
- **D6** : descripteur dual du site C8
- **D7** : descripteur dual du site S
- **D8** : descripteur dual du site O
- **D9** : Descripteur dual du site N12

Le tableau III-1 montre les neuf (09) descripteurs retenus pour construire les modèles QSAR. Il est important de souligner que les valeurs des descripteurs HEPT1, HEPT2, HEPT3, HEPT4, HEPT5, HEPT7, HEPT8, HEPT12, HEPT13, HEPT14, HEPT15, HEPT16, HEPT18, HEPT19, HEPT20, HEPT21, HEPT22, HEPT23, HEPT24, HEPT25, HEPT26, HEPT27, HEPT29, HEPT30, HEPT32, HEPT33, HEPT34, HEPT35, HEPT36, HEPT38, HEPT39, HEPT40, HEPT41, HEPT42, HEPT44, HEPT45, HEPT46, HEPT48, HEPT49, HEPT50, HEPT51, HEPT52, HEPT53, HEPT54, HEPT55, HEPT56, HEPT57, HEPT58, HEPT59, HEPT60, HEPT61, HEPT62, HEPT63, HEPT64, HEPT65, HEPT66, HEPT67, HEPT68, HEPT69, HEPT70, HEPT71, HEPT72, HEPT73, HEPT74, HEPT76, HEPT77, HEPT78, HEPT82, HEPT84, HEPT85, HEPT90, HEPT93, HEPT94, HEPT95, HEPT96, HEPT97, HEPT100, HEPT101, HEPT102, HEPT103, HEPT104, HEPT107 ont été repris des travaux de Touhami et al, (2019) [1].

Tableau III.1 Les valeurs des descripteurs.

N°	D1	D2	D3	D4	D5	D6	D7	D8	D9
1	7.64601	0.70746	0.1441223 31	2.514230676	0.817	0.643	0.30 4	- 0.581	- 0.623
2	8.65278	3.94545	0.2124346 5	8.429162558	0.838	0.654	0.10 7	- 0.535	- 0.598
3	8.02695	2.85705	0.1934273 39	5.728421053	0.83	0.676	- 0.13 3	- 0.663	- 0.621
4	7.83648	0.19047	0.1307871 69	2.106717304	0.844	0.658	0.20 63	- 0.718	- 0.641
5	7.961646	0.29931	0.1305085 03	2.226585797	0.838	0.652	0.07 7	- 0.694	- 0.639
6	7.939878	0.291147	0.1307406 42	2.214412187	0.84	0.654	0.07 5	- 0.695	- -0.64
7	8.05416	0.48978	0.1321985 41	2.412587374	0.849	0.667	0.12 5	- 0.727	- 0.631
8	8.201094	0.457128	0.1291327 98	2.420103865	0.841	0.657	0.09 7	- 0.722	- 0.637
9	8.179326	0.492501	0.1300927 24	2.445762181	0.841	0.658	0.09 9	- 0.691	- 0.637
10	8.168442	0.514269	0.1306476 87	2.462364984	0.84	0.658	0.10 3	-0.69	- 0.637
11	8.111301	0.557805	0.1323890 29	2.487371372	0.839	0.657	0.11 9	- 0.683	- 0.291

Chapitre 3 : Résultats et Discussions

12	8.163	1.03398	0.1402717 34	2.966201565	0.835	0.69	0.12	-	-
13	7.994298	0.304752	0.1300466 89	2.239216428	0.839	0.652	0.09 9	-	-
14	7.83648	0.8163	0.1424464 9	2.666263605	0.831	0.646	0.04 1	-0.65	-0.52
15	7.885458	0.266658	0.1312542 66	2.18069103	0.829	0.651	0.08	-	-
16	8.10858	1.0884	0.1424464 9	3.012189186	0.839	0.676	- 0.05 6	-	-
17	7.714035	0.500664	0.1386314 39	2.338756722	0.305	0.647	0.15 7	-	-
18	8.168442	0.702018	0.1339329 24	2.634630066	0.843	0.669	0.09 3	-	-
19	8.211978	0.848952	0.1358137 27	2.787592101	0.839	0.666	0.08 9	-	-
20	6.74808	3.42846	0.3012392 98	7.799233525	0.824	0.653	0.32 8	-	-
21	5.52363	1.68702	0.2606467 69	3.387982713	0.407	0.587	0.45 3	-	-0.64
22	7.83648	0.606783	0.1383183 83	2.465134088	0.308	0.659	0.13 9	-	-
23	7.828317	0.631272	0.1389459 15	2.4859038	0.307	0.644	0.13 3	-	-
24	7.798386	0.636714	0.1396321 98	2.483739552	0.309	0.651	0.13 7	-0.63	-
25	5.172621	-0.239448	0.1847722 19	1.124163229	2.816	3.569	0.44 4	-	-
26	7.681383	0.576852	0.1407552 45	2.399822216	0.31	0.649	0.13 5	-	-
27	5.610702	1.03398	0.2184969 94	2.411758835	-0.263	0.573	0.44 6	-	-
28	8.056881	0.378219	0.1302310 22	2.316513997	0.841	0.649	0.05 5	-	-
29	8.040555	0.394545	0.1307871 69	2.326406584	0.839	0.649	0.05 5	-	-
30	8.19021	0.46257	0.1294056 14	2.422169049	0.841	0.656	0.04	-	-
31	7.888179	0.312915	0.1320086 01	2.219656727	0.837	0.649	0.06 2	-	-
32	5.57805	0.68025	0.2041733 02	1.999179167	0.304	0.592	0.46 4	-	-0.75
33	8.08137	-3.997149	0.0827916 07	0.345258826	0.837	0.672	0.15 9	-	-
34	7.909947	0.269379	0.1308803 22	2.189018337	0.839	0.654	0.08 6	-	-0.64
35	8.048718	0.312915	0.1292690 62	2.259523233	0.841	0.656	0.08 3	-	-0.64
36	7.91811	1.00677	0.1446897 42	2.881260472	0.307	0.632	0.14 4	-	-
37	8.086812	0.658482	0.1346197 6	2.573935432	0.837	0.629	0.10 6	-	-
38	20.388453	21.427875	- 0.9620731 52	-420.5715502	0.87	0.69	0.21 9	-	-0.69
39	7.202487	1.292475	0.1692043 94	3.052632523	0.817	0.212	0.29 8	-	-
40	8.007903	0.527874	0.1336893 21	2.435133908	0.839	0.646	0.10 4	-	-
41	7.148067	0.770043	0.1567883 72	2.457519209	0.83	0.661	0.24 9	-	-
42	5.65968	0.68025	0.2008261 99	2.018037828	0.3	0.589	0.46	-	-
43	8.092254	0.375498	0.1295881 33	2.322958764	0.84	0.655	0.09 2	-	-

Chapitre 3 : Résultats et Discussions

44	7.986135	0.693855	0.1371313 22	2.582944785	0.837	0.659	0.11 1	- 0.675	- 0.635
45	5.60526	0.65304	0.2019296 4	1.977210165	0.293	0.8329	0.46	- 0.724	- 0.753
46	8.02695	0.5442	0.1336407 07	2.454465682	0.847	0.657	0.14 9	- 0.709	- 0.632
47	8.105859	0.514269	0.1317247 11	2.447004077	0.838	0.656	0.1	- 0.683	- 0.638
48	7.953483	0.906093	0.1418965 04	2.784438171	0.823	0.637	0.32 7	- 0.597	- 0.641
49	8.007903	0.280263	0.1294056 14	2.222337468	0.841	0.655	0.08 6	- 0.699	- 0.639
50	5.52363	0.73467	0.2088136 05	2.044615057	0.296	0.589	0.46	- 0.724	- 0.751
51	8.018787	0.334683	0.1301387 9	2.270286199	0.84	0.651	0.06 2	- 0.682	- 0.651
52	7.790223	0.568689	0.1384747 34	2.418856223	0.307	0.647	0.14 3	- 0.633	- 0.652
53	5.52363	0.57141	0.2019296 4	1.875396923	0.299	0.594	0.46 1	- 0.725	- 0.751
54	7.70043	0.46257	0.1381623 85	2.301597744	0.249	0.649	0.1	- 0.669	- 0.645
55	7.899063	0.394545	0.1332530 62	2.291417439	0.839	0.648	0.12 5	-0.67	- 0.638
56	7.768455	0.321078	0.1342754 64	2.196764853	0.835	0.646	0.11 5	-0.67	- 0.636
57	7.738524	0.612225	0.1403252 94	2.446396399	0.307	0.644	0.17 7	- 0.626	- 0.646
58	7.6188	0.541479	0.1412964 03	2.352237286	0.301	0.645	0.15 8	- 0.631	- 0.647
59	7.969809	0.356451	0.1313480 86	2.276479235	0.84	0.655	0.06	- 0.681	- -0.64
60	7.80927	0.48978	0.1366215 41	2.352425883	0.827	0.656	0.11 8	- 0.706	- 0.627
61	7.752129	0.590457	0.1396321 98	2.429556295	0.309	0.651	0.14 3	- 0.632	- -0.65
62	7.708593	0.631272	0.1412964 03	2.45690948	0.308	0.6333	0.17 5	- 0.626	- 0.647
63	8.048718	0.770043	0.1373876 43	2.671177981	0.844	0.669	0.07 4	- 0.703	- 0.639
64	7.99974	0.280263	0.1295424 55	2.220307466	0.84	0.654	0.08 3	- 0.697	- -0.64
65	7.828317	0.770043	0.1416776 96	2.618621588	0.828	0.673	0.11	- 0.714	- 0.634
66	7.97253	0.293868	0.1302310 22	2.224780043	0.835	0.656	0.10 4	-0.71	- 0.636
67	8.048718	6.941271	0.9029777 5	50.72472322	0.256	0.662	0.19 5	- 0.645	- 0.686
68	7.75485	0.612225	0.1400045 5	2.450357679	0.303	0.644	0.11 1	- 0.626	- 0.653
69	7.722198	0.5442	0.1393146 11	2.379958029	0.31	0.647	0.15 9	- 0.636	- -0.65
70	7.768455	0.595899	0.1394203 13	2.438545542	0.305	0.646	0.13 8	-0.63	- 0.652
71	7.629684	0.623109	0.1427230 85	2.430167104	0.344	0.824	0.18 9	-0.62	- 0.645
72	7.746687	0.731949	0.1425569 99	2.562008318	0.303	0.644	0.17 4	- 0.618	- 0.644
73	7.656894	0.57141	0.1411336 19	2.388862452	0.296	0.642	0.18 4	- 0.626	- 0.649
74	8.367075	1.069353	0.1370290 62	3.050478403	0.846	0.666	0.08 3	-0.67	- 0.642
75	7.942599	0.35373	0.1317719 41	2.267435203	0.839	0.649	0.04 5	- 0.678	- 0.642
76	7.958925	0.318357	0.1308803 22	2.241763875	0.842	0.652	0.07	- 0.687	- 0.637

Chapitre 3 : Résultats et Discussions

77	8.002461	0.359172	0.1308337 29	2.286872393	0.84	0.651	0.06 3	- 0.682	- 0.642
78	7.801107	0.397266	0.1350650 29	2.26954225	0.834	0.646	0.12 4	- 0.654	-0.64
79	7.599753	0.560526	0.1420610 53	2.364966827	0.32	0.655	0.19	- 0.635	- 0.648
80	7.765734	0.340125	0.1346690 89	2.212106446	0.839	0.654	0.12 6	- 0.677	- 0.635
81	7.986135	0.312915	0.1303233 84	2.243980718	0.839	0.65	0.11 4	- 0.684	-0.64
82	7.784781	1.164588	0.1510529 98	3.024504176	0.839	0.682	0.14	- 0.727	- 0.639
83	7.991577	0.315636	0.1302771 87	2.247600256	0.838	0.653	0.07 8	- 0.695	- 0.639
84	8.141232	0.394545	0.1290874 41	2.351311244	0.841	0.654	0.09 5	- 0.697	-0.64
85	8.094975	0.492501	0.1315361 29	2.42502454	0.84	0.655	0.10 4	- 0.688	- 0.639
86	8.505846	1.477503	0.1422810 47	3.545190426	0.842	0.672	0.11 9	- 0.683	-0.63
87	8.369796	0.95235	0.1348172 94	2.928986812	0.841	0.662	0.11 4	- 0.681	- 0.636
88	7.871853	0.242169	0.1310670 27	2.157276534	0.837	0.651	0.10 8	- 0.688	-0.64
89	7.752129	0.209517	0.1325800 66	2.100989917	0.835	0.648	0.11 5	- 0.682	-0.64
90	7.828317	1.371384	0.1548722 9	3.276884648	0.835	0.674	0.11 8	- 0.683	- 0.632
91	8.157558	0.737391	0.1347678 56	2.665712164	0.844	0.671	0.09 3	- 0.698	- 0.631
92	8.075928	0.555084	0.1329638	2.476263573	0.844	0.669	0.09 4	- 0.698	- 0.632
93	8.165721	0.497943	0.1304158 78	2.447223756	0.839	0.673	0.08 4	- 0.669	- 0.642
94	7.792944	0.601341	0.1390510 57	2.449524142	0.804	-0.115	0.14 8	- 0.626	- 0.723
95	7.539891	0.628551	0.1446897 42	2.413549496	0.832	0.632	0.28 1	- 0.603	- 0.644
96	13.03359	0.59862	0.0804183 69	3.736180093	0.826	0.671	0.37 4	- 0.579	-0.65
97	6.163065	1.662531	0.2221958 55	3.401815915	0.824	0.653	0.35 2	- 0.554	- 0.642
98	7.735803	0.242169	0.1334466 03	2.123403586	0.834	0.641	0.15 9	- 0.663	-0.64
99	8.40789	0.623109	0.1284557 65	2.619179105	0.841	0.655	0.05 6	- 0.652	- 0.646
100	7.684104	0.669366	0.1425569 99	2.486923284	0.898	0.575	- 0.05 1	- 0.721	- 0.402
101	6.497748	1.431246	0.1973748 36	3.102186965	0.827	0.651	0.37 7	-0.58	- 0.642
102	8.239188	-0.10884	0.1197887 69	1.979586035	0.841	0.604	0.10 3	- 0.671	- 0.643
103	8.198373	0.585015	0.1313480 86	2.533307404	0.013	0.067	- 0.11 9	- 0.032	- 0.002
104	8.045997	0.5442	0.1333013 94	2.459126943	0.837	0.633	0.11 1	- 0.648	- 0.641
105	7.86369	0.389103	0.1337866 56	2.278005203	0.837	0.649	0.14 4	- 0.672	- 0.637
106	8.296329	0.304752	0.1251317 48	2.314267709	0.82	0.657	0.07 5	- 0.693	-0.65
107	8.078649	0.310194	0.1287257 25	2.26469378	0.831	0.656	0.09 6	- 0.708	- 0.638

Chapitre 3 : Résultats et Discussions

Les structures chimiques ainsi que les données d'activité observées des composés utilisés dans cette étude sont indiquées dans le tableau III.2. Les données d'activité ont été rapportée à partir des travaux de Luco et Ferretti [2].

Tableau III.2 Structures chimiques et activités anti-VIH observées des dérivés HEPT

N°	R ₁	R ₂	R ₃	X	Activité
1	2-Me	Me	CH ₂ OCH ₂ CH ₂ OH	O	4.15
2	2-NO ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O	3.85
3	2-OMe	Me	CH ₂ OCH ₂ CH ₂ OH	O	4.72
4	3-Me	Me	CH ₂ OCH ₂ CH ₂ OH	O	5.59
5	3-Et	Me	CH ₂ OCH ₂ CH ₂ OH	O	5.57
6	3-t-Bu	Me	CH ₂ OCH ₂ CH ₂ OH	O	4.92
7	3-CF ₃	Me	CH ₂ OCH ₂ CH ₂ OH	O	4.35
8	3-F	Me	CH ₂ OCH ₂ CH ₂ OH	O	5.48
9	3-Cl	Me	CH ₂ OCH ₂ CH ₂ OH	O	4.89
10	3-Br	Me	CH ₂ OCH ₂ CH ₂ OH	O	5.24
11	3-I	Me	CH ₂ OCH ₂ CH ₂ OH	O	5.00
12	3-NO ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O	4.47
13	3-OH	Me	CH ₂ OCH ₂ CH ₂ OH	O	4.09
14	3-OMe	Me	CH ₂ OCH ₂ CH ₂ OH	O	4.66
15	3,5-Me ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O	6.59
16	3,5-Cl ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O	5.89
17	3,5-Me ₂	Me	CH ₂ OCH ₂ CH ₂ OH	S	6.66
18	3-COOMe	Me	CH ₂ OCH ₂ CH ₂ OH	O	5.10
19	3-COMe	Me	CH ₂ OCH ₂ CH ₂ OH	O	5.14
20	3-CN	Me	CH ₂ OCH ₂ CH ₂ OH	O	5.00
21	H	CH ₂ CH=CH ₂	CH ₂ OCH ₂ CH ₂ OH	O	5.60
22	H	Et	CH ₂ OCH ₂ CH ₂ OH	S	6.96
23	H	Pr	CH ₂ OCH ₂ CH ₂ OH	S	5.00
24	H	<i>i</i> -Pr	CH ₂ OCH ₂ CH ₂ OH	S	7.23
25	3,5-Me ₂	Et	CH ₂ OCH ₂ CH ₂ OH	S	8.11
26	3,5-Me ₂	<i>i</i> -Pr	CH ₂ OCH ₂ CH ₂ OH	S	8.30
27	3,5-Cl ₂	Et	CH ₂ OCH ₂ CH ₂ OH	S	7.37
28	H	Et	CH ₂ OCH ₂ CH ₂ OH	O	6.92
29	H	Pr	CH ₂ OCH ₂ CH ₂ OH	O	5.47
30	H	<i>i</i> -Pr	CH ₂ OCH ₂ CH ₂ OH	O	7.20
31	3,5-Me ₂	Et	CH ₂ OCH ₂ CH ₂ OH	O	7.89
32	3,5-Me ₂	<i>i</i> -Pr	CH ₂ OCH ₂ CH ₂ OH	O	8.57
33	3,5-Cl ₂	Et	CH ₂ OCH ₂ CH ₂ OH	O	7.85
34	4-Me	Me	CH ₂ OCH ₂ CH ₂ OH	O	3.66
35	H	Me	CH ₂ OCH ₂ CH ₂ OH	O	5.15
36	H	Me	CH ₂ OCH ₂ CH ₂ OH	S	6.01
37	H	I	CH ₂ OCH ₂ CH ₂ OH	O	5.44
38	H	CH=CH ₂	CH ₂ OCH ₂ CH ₂ OH	O	5.69
39	H	CH=CHPh	CH ₂ OCH ₂ CH ₂ OH	O	5.22
40	H	CH ₂ Ph	CH ₂ OCH ₂ CH ₂ OH	O	4.37

Chapitre 3 : Résultats et Discussions

41	H	CH=CPh ₂	CH ₂ OCH ₂ CH ₂ OH	O	6.07
42	H	Me	CH ₂ OCH ₂ CH ₂ OMe	O	5.06
43	H	Me	CH ₂ OCH ₂ CH ₂ OAc	O	5.17
44	H	Me	CH ₂ OCH ₂ CH ₂ OCOPh	O	5.12
45	H	Me	CH ₂ OCH ₂ Me	O	6.48
46	H	Me	CH ₂ OCH ₂ CH ₂ Cl	O	5.82
47	H	Me	CH ₂ OCH ₂ CH ₂ N ₃	O	5.24
48	H	Me	CH ₂ OCH ₂ CH ₂ F	O	5.96
49	H	Me	CH ₂ OCH ₂ CH ₂ Me	O	5.48
50	H	Me	CH ₂ OCH ₂ Ph	O	7.06
51	H	Et	CH ₂ OCH ₂ Me	O	7.72
52	H	Et	CH ₂ OCH ₂ Me	S	7.58
53	3,5-Me ₂	Et	CH ₂ OCH ₂ Me	O	8.24
54	3,5-Me ₂	Et	CH ₂ OCH ₂ Me	S	8.30
55	H	Et	CH ₂ OCH ₂ Ph	O	8.23
56	3,5-Me ₂	Et	CH ₂ OCH ₂ Ph	O	8.55
57	H	Et	CH ₂ OCH ₂ Ph	S	8.09
58	3,5-Me ₂	Et	CH ₂ OCH ₂ Ph	S	8.14
59	H	<i>i</i> -Pr	CH ₂ OCH ₂ Me	O	7.99
60	H	<i>i</i> -Pr	CH ₂ OCH ₂ Ph	O	8.51
61	H	<i>i</i> -Pr	CH ₂ OCH ₂ Me	S	7.89
62	H	<i>i</i> -Pr	CH ₂ OCH ₂ Ph	S	8.14
63	H	Me	CH ₂ OMe	O	5.68
64	H	Me	CH ₂ OBu	O	5.33
65	H	Me	Et	O	5.66
66	H	Me	Bu	O	5.92
67	3,5-Cl ₂	Et	CH ₂ OCH ₂ Me	S	7.89
68	H	Et	CH ₂ O- <i>i</i> -Pr	S	6.66
69	H	Et	CH ₂ O- <i>c</i> -Hex	S	5.79
70	H	Et	CH ₂ OCH ₂ - <i>c</i> -Hex	S	6.45
71	H	Et	CH ₂ OCH ₂ C ₆ H ₄ (4-Me)	S	7.11
72	H	Et	CH ₂ OCH ₂ C ₆ H ₄ (4-Cl)	S	7.92
73	H	Et	CH ₂ OCH ₂ CH ₂ Ph	S	7.04
74	3,5-Cl ₂	Et	CH ₂ OCH ₂ Me	O	8.13
75	H	Et	CH ₂ O- <i>i</i> -Pr	O	6.47
76	H	Et	CH ₂ O- <i>c</i> -Hex	O	5.40
77	H	Et	CH ₂ OCH ₂ - <i>c</i> -Hex	O	6.35
78	H	Et	CH ₂ OCH ₂ CH ₂ Ph	O	7.02
79	H	<i>c</i> -Pr	CH ₂ OCH ₂ Me	S	7.02
80	H	<i>c</i> -Pr	CH ₂ OCH ₂ Me	O	7.00
81	H	Me	CH ₂ OCH ₂ CH ₂ OC ₅ H _{11-<i>n</i>}	O	<4.46
82	2-Cl	Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.89
83		Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.53
84	4-F	Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.60
85	4-Cl	Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.60
86	4-NO ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.72

87	4-CN	Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.60
88	4-OH	Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.56
89	4-OMe	Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.60
90	4-COMe	Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.96
91	3-COOH	Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.45
92	3-CONH ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.51
93	H	COOMe	CH ₂ OCH ₂ CH ₂ OH	O	<5.18
94	H	CONHPh	CH ₂ OCH ₂ CH ₂ OH	O	<4.74
95	H	SPh	CH ₂ OCH ₂ CH ₂ OH	O	<4.68
96	H	C≡CH	CH ₂ OCH ₂ CH ₂ OH	O	<4.74
97	H	C≡C-Ph	CH ₂ OCH ₂ CH ₂ OH	O	<5.47
98	3-NH ₂	Me	CH ₂ OCH ₂ CH ₂ OH	O	<3.60
99	H	COCHMe ₂	CH ₂ OCH ₂ CH ₂ OH	O	<4.92
100	H	COPh	CH ₂ OCH ₂ CH ₂ OH	O	<4.89
101	H	C≡CMe	CH ₂ OCH ₂ CH ₂ OH	O	<4.72
102	H	F	CH ₂ OCH ₂ CH ₂ OH	O	<4.00
103	H	Cl	CH ₂ OCH ₂ CH ₂ OH	O	<4.52
104	H	Br	CH ₂ OCH ₂ CH ₂ OH	O	<4.70
105	H	Me	CH ₂ OCH ₂ CH ₂ OCH ₂ Ph	O	<4.70
106	H	Me	H	O	<3.60
107	H	Me	Me	O	<3.82

III.3 Sélection des données d'apprentissage et de test

La division des données en ensembles d'apprentissage et de test doit être exécutée de telle manière que les points représentant à la fois les ensembles d'apprentissage et de test soient dispersés dans tout l'espace descripteur occupé par l'ensemble de données entier et chaque point de l'ensemble de test est près d'au moins un composé de l'ensemble d'apprentissage. Dans notre travail le logiciel « *Dataset Division 1.2* » a été utilisé pour diviser l'ensemble des données avec la méthode 'Activity/Property based' [3].

Les tableaux III.3 et III.4 présentent l'ensemble des données d'apprentissage et de test sélectionnées par l'outil **Dataset Division 1.2**

Tableau III.3 Les données d'apprentissage (training).

N°	D1	D2	D3	D4	D5	D6	D7	D8	D9
1	7.64601	0.7074 6	0.1441223 31	2.5142 30676	0.8 17	0.64 3	0.30 4	-0.581	-0.623
2	8.65278	3.9454 5	0.2124346 5	8.4291 62558	0.8 38	0.65 4	0.10 7	-0.535	-0.598
3	8.02695	2.8570 5	0.1934273 39	5.7284 21053	0.8 3	0.67 6	- 0.13 3	-0.663	-0.621
5	7.96164 6	0.2993 1	0.1305085 03	2.2265 85797	0.8 38	0.65 2	0.07 7	-0.694	-0.639

Chapitre 3 : Résultats et Discussions

7	8.05416	0.4897 8	0.1321985 41	2.4125 87374	0.8 49	0.66 7	0.12 5	-0.727	-0.631
8	8.20109 4	0.4571 28	0.1291327 98	2.4201 03865	0.8 41	0.65 7	0.09 7	-0.722	-0.637
9	8.17932 6	0.4925 01	0.1300927 24	2.4457 62181	0.8 41	0.65 8	0.09 9	-0.691	-0.637
11	8.11130 1	0.5578 05	0.1323890 29	2.4873 71372	0.8 39	0.65 7	0.11 9	-0.683	-0.291
12	8.163	1.0339 8	0.1402717 34	2.9662 01565	0.8 35	0.69	0.12	-0.691	-0.624
13	7.99429 8	0.3047 52	0.1300466 89	2.2392 16428	0.8 39	0.65 2	0.09 9	-0.686	-0.638
14	7.83648	0.8163	0.1424464 9	2.6662 63605	0.8 31	0.64 6	0.04 1	-0.65	-0.52
15	7.88545 8	0.2666 58	0.1312542 66	2.1806 9103	0.8 29	0.65 1	0.08	-0.691	-0.638
16	8.10858	1.0884	0.1424464 9	3.0121 89186	0.8 39	0.67 6	- 0.05 6	-0.699	-0.637
17	7.71403 5	0.5006 64	0.1386314 39	2.3387 56722	0.3 05	0.64 7	0.15 7	-0.638	-0.653
18	8.16844 2	0.7020 18	0.1339329 24	2.6346 30066	0.8 43	0.66 9	0.09 3	-0.698	-0.632
21	5.52363	1.6870 2	0.2606467 69	3.3879 82713	0.4 07	0.58 7	0.45 3	-0.703	-0.64
22	7.83648	0.6067 83	0.1383183 83	2.4651 34088	0.3 08	0.65 9	0.13 9	-0.631	-0.654
23	7.82831 7	0.6312 72	0.1389459 15	2.4859 038	0.3 07	0.64 4	0.13 3	-0.628	-0.653
25	5.17262 1	- 0.2394 48	0.1847722 19	1.1241 63229	2.8 16	3.56 9	0.44 4	-0.823	-0.675
26	7.68138 3	0.5768 52	0.1407552 45	2.3998 22216	0.3 1	0.64 9	0.13 5	-0.631	-0.649
27	5.61070 2	1.0339 8	0.2184969 94	2.4117 58835	- 0.2 63	0.57 3	0.44 6	-0.779	-0.856
28	8.05688 1	0.3782 19	0.1302310 22	2.3165 13997	0.8 41	0.64 9	0.05 5	-0.679	-0.641
30	8.19021	0.4625 7	0.1294056 14	2.4221 69049	0.8 41	0.65 6	0.04	-0.715	-0.631
32	5.57805	0.6802 5	0.2041733 02	1.9991 79167	0.3 04	0.59 2	0.46 4	-0.733	-0.75
33	8.08137	- 3.9971 49	0.0827916 07	0.3452 58826	0.8 37	0.67 2	0.15 9	-0.672	-0.635
34	7.90994 7	0.2693 79	0.1308803 22	2.1890 18337	0.8 39	0.65 4	0.08 6	-0.694	-0.64
35	8.04871 8	0.3129 15	0.1292690 62	2.2595 23233	0.8 41	0.65 6	0.08 3	-0.698	-0.64
37	8.08681 2	0.6584 82	0.1346197 6	2.5739 35432	0.8 37	0.62 9	0.10 6	-0.642	-0.641
38	20.3884 53	21.427 875	- 0.9620731 52	- 420.57 15502	0.8 7	0.69	0.21 9	-0.567	-0.69
39	7.20248 7	1.2924 75	0.1692043 94	3.0526 32523	0.8 17	0.21 2	0.29 8	-0.621	-0.673
41	7.14806 7	0.7700 43	0.1567883 72	2.4575 19209	0.8 3	0.66 1	0.24 9	-0.589	-0.634

Chapitre 3 : Résultats et Discussions

42	5.65968	0.6802 5	0.2008261 99	2.0180 37828	0.3	0.58 9	0.46	-0.723	-0.752
43	8.09225 4	0.3754 98	0.1295881 33	2.3229 58764	0.8 4	0.65 5	0.09 2	-0.692	-0.639
44	7.98613 5	0.6938 55	0.1371313 22	2.5829 44785	0.8 37	0.65 9	0.11 1	-0.675	-0.635
45	5.60526	0.6530 4	0.2019296 4	1.9772 10165	0.2 93	0.83 29	0.46	-0.724	-0.753
46	8.02695	0.5442	0.1336407 07	2.4544 65682	0.8 47	0.65 7	0.14 9	-0.709	-0.632
47	8.10585 9	0.5142 69	0.1317247 11	2.4470 04077	0.8 38	0.65 6	0.1	-0.683	-0.638
48	7.95348 3	0.9060 93	0.1418965 04	2.7844 38171	0.8 23	0.63 7	0.32 7	-0.597	-0.641
49	8.00790 3	0.2802 63	0.1294056 14	2.2223 37468	0.8 41	0.65 5	0.08 6	-0.699	-0.639
50	5.52363	0.7346 7	0.2088136 05	2.0446 15057	0.2 96	0.58 9	0.46	-0.724	-0.751
51	8.01878 7	0.3346 83	0.1301387 9	2.2702 86199	0.8 4	0.65 1	0.06 2	-0.682	-0.651
52	7.79022 3	0.5686 89	0.1384747 34	2.4188 56223	0.3 07	0.64 7	0.14 3	-0.633	-0.652
53	5.52363	0.5714 1	0.2019296 4	1.8753 96923	0.2 99	0.59 4	0.46 1	-0.725	-0.751
54	7.70043	0.4625 7	0.1381623 85	2.3015 97744	0.2 49	0.64 9	0.1	-0.669	-0.645
56	7.76845 5	0.3210 78	0.1342754 64	2.1967 64853	0.8 35	0.64 6	0.11 5	-0.67	-0.636
58	7.6188	0.5414 79	0.1412964 03	2.3522 37286	0.3 01	0.64 5	0.15 8	-0.631	-0.647
59	7.96980 9	0.3564 51	0.1313480 86	2.2764 79235	0.8 4	0.65 5	0.06	-0.681	-0.64
60	7.80927	0.4897 8	0.1366215 41	2.3524 25883	0.8 27	0.65 6	0.11 8	-0.706	-0.627
61	7.75212 9	0.5904 57	0.1396321 98	2.4295 56295	0.3 09	0.65 1	0.14 3	-0.632	-0.65
62	7.70859 3	0.6312 72	0.1412964 03	2.4569 0948	0.3 08	0.63 33	0.17 5	-0.626	-0.647
63	8.04871 8	0.7700 43	0.1373876 43	2.6711 77981	0.8 44	0.66 9	0.07 4	-0.703	-0.639
64	7.99974	0.2802 63	0.1295424 55	2.2203 07466	0.8 4	0.65 4	0.08 3	-0.697	-0.64
65	7.82831 7	0.7700 43	0.1416776 96	2.6186 21588	0.8 28	0.67 3	0.11	-0.714	-0.634
66	7.97253	0.2938 68	0.1302310 22	2.2247 80043	0.8 35	0.65 6	0.10 4	-0.71	-0.636
67	8.04871 8	6.9412 71	0.9029777 5	50.724 72322	0.2 56	0.66 2	0.19 5	-0.645	-0.686
68	7.75485	0.6122 25	0.1400045 5	2.4503 57679	0.3 03	0.64 4	0.11 1	-0.626	-0.653
70	7.76845 5	0.5958 99	0.1394203 13	2.4385 45542	0.3 05	0.64 6	0.13 8	-0.63	-0.652
71	7.62968 4	0.6231 09	0.1427230 85	2.4301 67104	0.3 44	0.82 4	0.18 9	-0.62	-0.645
72	7.74668 7	0.7319 49	0.1425569 99	2.5620 08318	0.3 03	0.64 4	0.17 4	-0.618	-0.644
73	7.65689 4	0.5714 1	0.1411336 19	2.3888 62452	0.2 96	0.64 2	0.18 4	-0.626	-0.649
74	8.36707 5	1.0693 53	0.1370290 62	3.0504 78403	0.8 46	0.66 6	0.08 3	-0.67	-0.642

Chapitre 3 : Résultats et Discussions

76	7.95892 5	0.3183 57	0.1308803 22	2.2417 63875	0.8 42	0.65 2	0.07	-0.687	-0.637
77	8.00246 1	0.3591 72	0.1308337 29	2.2868 72393	0.8 4	0.65 1	0.06 3	-0.682	-0.642
78	7.80110 7	0.3972 66	0.1350650 29	2.2695 4225	0.8 34	0.64 6	0.12 4	-0.654	-0.64
80	7.76573 4	0.3401 25	0.1346690 89	2.2121 06446	0.8 39	0.65 4	0.12 6	-0.677	-0.635
81	7.98613 5	0.3129 15	0.1303233 84	2.2439 80718	0.8 39	0.65	0.11 4	-0.684	-0.64
82	7.78478 1	1.1645 88	0.1510529 98	3.0245 04176	0.8 39	0.68 2	0.14	-0.727	-0.639
83	7.99157 7	0.3156 36	0.1302771 87	2.2476 00256	0.8 38	0.65 3	0.07 8	-0.695	-0.639
84	8.14123 2	0.3945 45	0.1290874 41	2.3513 11244	0.8 41	0.65 4	0.09 5	-0.697	-0.64
87	8.36979 6	0.9523 5	0.1348172 94	2.9289 86812	0.8 41	0.66 2	0.11 4	-0.681	-0.636
88	7.87185 3	0.2421 69	0.1310670 27	2.1572 76534	0.8 37	0.65 1	0.10 8	-0.688	-0.64
89	7.75212 9	0.2095 17	0.1325800 66	2.1009 89917	0.8 35	0.64 8	0.11 5	-0.682	-0.64
92	8.07592 8	0.5550 84	0.1329638	2.4762 63573	0.8 44	0.66 9	0.09 4	-0.698	-0.632
93	8.16572 1	0.4979 43	0.1304158 78	2.4472 23756	0.8 39	0.67 3	0.08 4	-0.669	-0.642
94	7.79294 4	0.6013 41	0.1390510 57	2.4495 24142	0.8 04	- 0.11 5	0.14 8	-0.626	-0.723
95	7.53989 1	0.6285 51	0.1446897 42	2.4135 49496	0.8 32	0.63 2	0.28 1	-0.603	-0.644
97	6.16306 5	1.6625 31	0.2221958 55	3.4018 15915	0.8 24	0.65 3	0.35 2	-0.554	-0.642
98	7.73580 3	0.2421 69	0.1334466 03	2.1234 03586	0.8 34	0.64 1	0.15 9	-0.663	-0.64
99	8.40789	0.6231 09	0.1284557 65	2.6191 79105	0.8 41	0.65 5	0.05 6	-0.652	-0.646
100	7.68410 4	0.6693 66	0.1425569 99	2.4869 23284	0.8 98	0.57 5	- 0.05 1	-0.721	-0.402
101	6.49774 8	1.4312 46	0.1973748 36	3.1021 86965	0.8 27	0.65 1	0.37 7	-0.58	-0.642
102	8.23918 8	- 0.1088 4	0.1197887 69	1.9795 86035	0.8 41	0.60 4	0.10 3	-0.671	-0.643
103	8.19837 3	0.5850 15	0.1313480 86	2.5333 07404	0.0 13	0.06 7	- 0.11 9	-0.032	-0.002
105	7.86369	0.3891 03	0.1337866 56	2.2780 05203	0.8 37	0.64 9	0.14 4	-0.672	-0.637
106	8.29632 9	0.3047 52	0.1251317 48	2.3142 67709	0.8 2	0.65 7	0.07 5	-0.693	-0.65
107	8.07864 9	0.3101 94	0.1287257 25	2.2646 9378	0.8 31	0.65 6	0.09 6	-0.708	-0.638

Tableau III.4 Les données de test.

N°	D1	D2	D3	D4	D5	D6	D7	D8	D9
4	7.83648	0.19047	0.130787 169	2.10671 7304	0.844	0.65 8	0.206 3	-0.718	- 0.641
6	7.939878	0.291147	0.130740 642	2.21441 2187	0.84	0.65 4	0.075	-0.695	-0.64
10	8.168442	0.514269	0.130647 687	2.46236 4984	0.84	0.65 8	0.103	-0.69	- 0.637
19	8.211978	0.848952	0.135813 727	2.78759 2101	0.839	0.66 6	0.089	-0.689	- 0.633
20	6.74808	3.42846	0.301239 298	7.79923 3525	0.824	0.65 3	0.328	-0.572	- 0.631
24	7.798386	0.636714	0.139632 198	2.48373 9552	0.309	0.65 1	0.137	-0.63	- 0.649
29	8.040555	0.394545	0.130787 169	2.32640 6584	0.839	0.64 9	0.055	-0.675	- 0.643
31	7.888179	0.312915	0.132008 601	2.21965 6727	0.837	0.64 9	0.062	-0.679	- 0.642
36	7.91811	1.00677	0.144689 742	2.88126 0472	0.307	0.63 2	0.144	-0.662	- 0.646
40	8.007903	0.527874	0.133689 321	2.43513 3908	0.839	0.64 6	0.104	-0.657	- 0.641
55	7.899063	0.394545	0.133253 062	2.29141 7439	0.839	0.64 8	0.125	-0.67	- 0.638
57	7.738524	0.612225	0.140325 294	2.44639 6399	0.307	0.64 4	0.177	-0.626	- 0.646
69	7.722198	0.5442	0.139314 611	2.37995 8029	0.31	0.64 7	0.159	-0.636	-0.65
75	7.942599	0.35373	0.131771 941	2.26743 5203	0.839	0.64 9	0.045	-0.678	- 0.642
79	7.599753	0.560526	0.142061 053	2.36496 6827	0.32	0.65 5	0.19	-0.635	- 0.648
85	8.094975	0.492501	0.131536 129	2.42502 454	0.84	0.65 5	0.104	-0.688	- 0.639
86	8.505846	1.477503	0.142281 047	3.54519 0426	0.842	0.67 2	0.119	-0.683	-0.63
90	7.828317	1.371384	0.154872 29	3.27688 4648	0.835	0.67 4	0.118	-0.683	- 0.632
91	8.157558	0.737391	0.134767 856	2.66571 2164	0.844	0.67 1	0.093	-0.698	- 0.631
96	13.03359	0.59862	0.080418 369	3.73618 0093	0.826	0.67 1	0.374	-0.579	-0.65
104	8.045997	0.5442	0.133301 394	2.45912 6943	0.837	0.63 3	0.111	-0.648	- 0.641

III.4 Etude quantitative des relations structure-activité QSAR

De nombreux modèles de relation structure-activité des composés HEPT ont été développés au cours des années avec des méthodes statistiques et un ensemble de descripteurs différents. Le tableau III.5 représente une comparaison de différentes méthodes d'analyse QSAR de composés HEPT.

Tableau III.5 Comparaison de différentes méthodes d'analyse QSAR de dérivés HEPT [1] avec notre travail.

Référence	Méthode	Nombre molécule	Nombre de descripteurs	R ²	Type de descripteurs
[4]	MLR	33	5	0.941	Hydrophobiques
[5]	MLR	40	11	0.869	Electroniques
[2]	MLR	107	10	0.951	Hydrophobiques et géométriques
	PLS	107	10	0.944	
[6]	MLR	87	8	0.940	Flexibilité conformationnelle des substituant, physico-chimiques, géométriques et électroniques
[7]	MLR	107	6	0.901	Topologiques géométriques, Électroniques, Et physicochimiques
	RNA	107	6	/	
[8]	MLR	76	4	0.960	Physico-chimiques, électroniques et géométriques
[9]	RNA	80	8	0.984	Hydrophobiques et géométriques
[10]	PSO	132	40	0.900	Hydrophobiques, longueur de substituant et indices de connectivité
[11]	MLR	79	6	0.840	Stériques, hydrophobiques, électroniques et structurales
	RNA	79	6	0.862	
	SVM	79	6	0.880	
[12]	RF	132	5	0.830	Hydrophobiques en relation avec la surface de van der Waals et le nombre d'atomes d'oxygène
[13]	MLR	34	9	0.919	Physico-chimiques et quantiques
[1]	MLR	60	9	0.806	Indices de Fukui

Chapitre 3 : Résultats et Discussions

Dans ce travail, deux méthodes statistiques sont utilisées dans le développement des modèles QSAR qui sont : la régression multi linéaire et réseaux de neurones artificiels (MLP perceptron multi linéaire).

L'analyse statistique des descripteurs moléculaires a été réalisée au moyen de logiciel Weka pour Windows [14]. En premier temps, En utilisant la méthode MLP (perceptron multi linéaire), un modèle QSAR a été développé sur le même ensemble de 60 dérivés HEPT utilisés dans les travaux de Touhami et al, (2019) [1]. Les résultats sont reportés dans le tableau III-6.

Tableau III.6 Performance du MLP.

Référence	Méthode	Nombre molécule	Nombre de descripteurs	R ²	Type de descripteurs
[1]	MLR	60	9	0.806	Indices de Fukui
Notre travail	MLP	60	9	0.864	Indices de Fukui

Le tableau III.7 regroupe les valeurs des activités observées et prédites en utilisant la méthode MLP sur 60 dérivés HEPT.

Tableau III.7 Les valeurs des activités observées/calculées en utilisant le MLP.

N°	Observée	Calculée	Résiduel
1	4.15	4.733	0.583
2	3.85	3.884	0.034
3	4.72	4.853	0.133
4	5.59	4.927	-0.663
5	5.57	5.161	-0.409
7	4.35	4.812	0.462
8	5.48	4.957	-0.523
12	4.47	4.805	0.335
13	4.09	5.105	1.015
14	4.66	3.744	-0.916
15	6.59	5.196	-1.394
16	5.89	5.307	-0.583
18	5.1	4.926	-0.174
19	5.14	4.938	-0.202
20	5	4.536	-0.464
21	5.6	5.904	0.304
22	6.96	7.311	0.351
24	7.23	7.258	0.028
25	8.11	8.279	0.169
26	8.3	7.264	-1.036
29	5.47	5.285	-0.185
32	8.57	7.556	-1.014

35	5.15	5.148	-0.002
36	6.01	7.097	1.087
38	5.69	5.436	-0.254
40	4.37	5.145	0.775
41	6.07	5.061	-1.009
44	5.12	5.009	-0.111
46	5.82	4.785	-1.035
48	5.96	4.823	-1.137
49	5.48	5.128	-0.352
50	7.06	7.57	0.51
51	7.72	5.438	-2.282
52	7.58	7.275	-0.305
53	8.24	7.576	-0.664
54	8.3	7.334	-0.966
57	8.09	7.192	-0.898
58	8.14	7.241	-0.899
61	7.89	7.264	-0.626
62	8.14	7.177	-0.963
63	5.68	5.082	-0.598
64	5.33	5.16	-0.17
65	5.66	5.009	-0.651
66	5.92	5.041	-0.879
67	7.89	7.074	-0.816
68	6.66	7.316	0.656
70	6.45	7.283	0.833
71	7.11	7.448	0.338
72	7.92	7.184	-0.736
73	7.04	7.238	0.198
76	5.4	5.132	-0.268
77	6.35	5.248	-1.102
90	3.96	4.878	0.918
93	5.18	5.293	0.113
95	4.68	4.989	0.309
96	4.74	4.321	-0.419
97	5.47	5.05	-0.42
101	4.72	4.925	0.205
103	4.52	4.991	0.471
104	4.7	5.091	0.391

Ensuite, deux modèles QSAR ont été développés sur l'ensemble des 107 dérivés HEPT, où 86 dérivés ont été utilisés pour apprentissage et validation interne et 21 dérivés ont été utilisés pour la validation externe des modèles.

1. La régression linéaire

L'analyse de la régression de l'activité étudiée nous a donné le modèle statistique comme suit :

$$\text{Log1/EC50} = 0.1515 * D1 + (-0.1004) * D2 + (-2.9801) * D5 + 2.8372 * D6 + (-2.5147) * D9 + 3.3591$$

Chapitre 3 : Résultats et Discussions

Les résultats obtenus sont indiqués dans le Tableau 8.

Tableau III.8 Performance de la régression linéaire sur les données d'apprentissage et les données de test pour les dérivés HEPT.

	R²
Evaluation Interne (apprentissage)	0.5611
Evaluation Externe (Test)	0.6359

Tableau III.9 Les valeurs expérimentales, prédites et résiduelles des activités par MLR sur les données d'apprentissage.

N°	Observée	Calculée	Résiduel
1	4.15	5.402	1.252
2	3.85	5.135	1.285
3	4.72	5.294	0.574
5	5.57	5.494	-0.076
7	4.35	5.479	1.129
8	5.48	5.515	0.035
9	4.89	5.511	0.621
11	5	4.627	-0.373
12	4.47	5.53	1.06
13	4.09	5.493	1.403
14	4.66	5.128	0.468
15	6.59	5.508	-1.082
16	5.89	5.497	-0.393
17	6.66	7.046	0.386
18	5.1	5.501	0.401
21	5.6	6.088	0.488
22	6.96	7.082	0.122
23	5	7.036	2.036
25	8.11	7.598	-0.512
26	8.3	7.014	-1.286
27	7.37	8.667	1.297
28	6.92	5.488	-1.432
30	7.2	5.495	-1.705
32	8.57	6.795	-1.775
33	7.85	5.994	-1.856
34	3.66	5.495	1.835
35	5.15	5.511	0.361
37	5.44	5.42	-0.02
38	5.69	5.395	-0.295
39	5.22	4.179	-1.041
41	6.07	5.361	-0.709

Chapitre 3 : Résultats et Discussions

42	5.06	6.816	1.756
43	5.17	5.509	0.339
44	5.12	5.471	0.351
45	6.48	7.526	1.046
46	5.82	5.449	-0.371
47	5.24	5.503	0.263
48	5.96	5.439	-0.521
49	5.48	5.503	0.023
50	7.06	6.8	-0.26
51	7.72	5.521	-2.199
52	7.58	7.042	-0.538
53	8.24	6.821	-1.419
54	8.3	7.2	-1.1
56	8.55	5.447	-3.103
58	8.14	7.019	-1.121
59	7.99	5.495	-2.495
60	8.51	5.466	-3.044
61	7.89	7.035	-0.855
62	8.14	6.969	-1.171
63	5.68	5.491	-0.189
64	5.33	5.504	0.174
65	5.66	5.504	-0.156
66	5.92	5.509	-0.411
67	7.89	6.721	-1.169
68	6.66	7.039	0.379
70	6.45	7.039	0.589
71	7.11	7.387	0.277
72	7.92	7.003	-0.917
73	7.04	7.033	-0.007
74	8.13	5.502	-2.628
76	5.4	5.475	0.075
77	6.35	5.493	-0.857
78	7.02	5.458	-1.562
80	7	5.453	-1.547
81	6.46	5.491	-0.969
82	3.89	5.463	1.573
83	3.53	5.5	1.97
84	3.6	5.511	1.911
87	3.6	5.503	1.903
88	3.56	5.489	1.929
89	3.6	5.472	1.872
92	3.51	5.499	1.989
93	5.18	5.569	0.389

94	4.74	3.575	-1.165
95	4.68	5.371	0.691
97	5.47	5.137	-0.333
98	3.6	5.449	1.849
99	4.92	5.547	0.627
100	4.89	4.422	-0.468
101	4.72	5.196	0.476
102	4	5.442	1.442
103	4.52	4.698	0.178
105	4.7	5.46	0.76
106	3.82	5.64	1.82
107	3.82	5.541	1.721

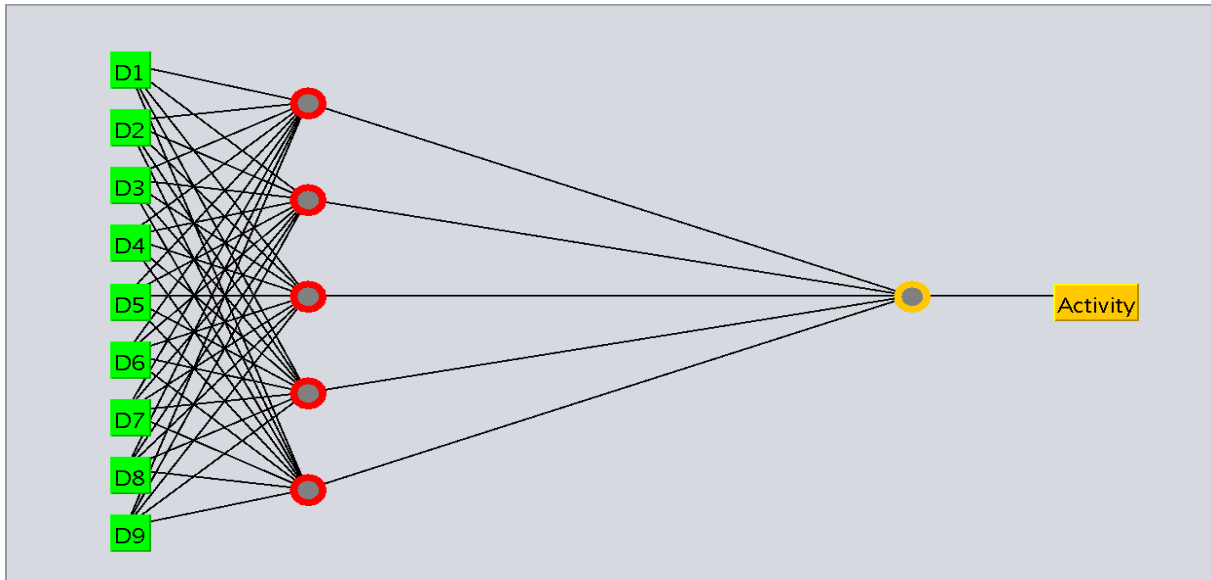
Tableau III.10 Les valeurs expérimentales, prédites et résiduelles des activités par MLR sur les données de test.

N°	Observée	Calculée	Résiduel
4	5.59	5.882	0.292
6	6.92	5.685	-1.235
10	5.24	5.134	-0.106
19	5.14	4.532	-0.608
20	5	4.975	-0.025
24	7.23	7.063	-0.167
29	5.47	6.07	0.6
31	7.89	6.001	-1.889
36	6.01	6.567	0.557
40	4.37	5.796	1.426
55	8.23	5.415	-2.815
57	8.09	6.657	-1.433
69	5.79	7.246	1.456
75	6.47	5.974	-0.496
79	7.02	7.017	-0.003
85	3.6	5.462	1.862
86	3.72	3.948	0.228
90	3.96	4.579	0.619
91	3.45	4.268	0.818
96	4.74	4.582	-0.158
104	4.7	5.777	1.077

2. Perceptron multicouche

La Figure 1 montre les couches du perceptron obtenues par le MLP pour les dérivés Hept.

Figure III.1 Perceptron multicouche du modèle développé pour les dérivés HEPT.



Le Tableau III.11 montre les résultats du modèle basé sur MLP comme suit :

Tableau III.11 Performance du perceptron multi couches sur les données d'apprentissage et les données de test pour les dérivés HEPT.

	R²
Evaluation Interne (apprentissage)	0.5826
Evaluation Externe (Test)	0.4933

Tableau III.12 Les valeurs expérimentales, prédites et résiduelles des activités par MLP sur les données d'apprentissage.

N°	Observée	Calculée	Résiduel
1	4.15	6.099	1.949
2	3.85	5.102	1.252
3	4.72	5.565	0.845
5	5.57	5.88	0.31
7	4.35	5.849	1.499
8	5.48	5.831	0.351
9	4.89	5.849	0.959
11	5	7.276	2.276
12	4.47	5.865	1.395
13	4.09	5.889	1.799
14	4.66	6.092	1.432
15	6.59	5.915	-0.675

16	5.89	5.765	-0.125
17	6.66	8.111	1.451
18	5.1	5.832	0.732
21	5.6	7.317	1.717
22	6.96	8.11	1.15
23	5	8.021	3.021
25	8.11	9.536	1.426
26	8.3	8.081	-0.219
27	7.37	8.359	0.989
28	6.92	5.847	-1.073
30	7.2	5.829	-1.371
32	8.57	7.03	-1.54
33	7.85	7.619	-0.231
34	3.66	5.892	2.232
35	5.15	5.873	0.723
37	5.44	5.792	0.352
38	5.69	6.756	1.066
39	5.22	5.28	0.06
41	6.07	6.05	-0.02
42	5.06	7.035	1.975
43	5.17	5.867	0.697
44	5.12	5.853	0.733
45	6.48	8.456	1.976
46	5.82	5.839	0.019
47	5.24	5.857	0.617
48	5.96	5.919	-0.041
49	5.48	5.88	0.4
50	7.06	7.047	-0.013
51	7.72	5.851	-1.869
52	7.58	8.076	0.496
53	8.24	7.099	-1.141
54	8.3	8.299	-0.001
56	8.55	5.914	-2.636
58	8.14	8.178	0.038
59	7.99	5.873	-2.117
60	8.51	5.914	-2.596
61	7.89	8.1	0.21
62	8.14	8.087	-0.053
63	5.68	5.81	0.13
64	5.33	5.88	0.55
65	5.66	5.879	0.219
66	5.92	5.899	-0.021
67	7.89	8.364	0.474

Chapitre 3 : Résultats et Discussions

68	6.66	8.009	1.349
70	6.45	8.067	1.617
71	7.11	8.998	1.888
72	7.92	8.189	0.269
73	7.04	8.221	1.181
74	8.13	5.739	-2.391
76	5.4	5.873	0.473
77	6.35	5.86	-0.49
78	7.02	5.907	-1.113
80	7	5.92	-1.08
81	6.46	5.885	-0.575
82	3.89	5.8	1.91
83	3.53	5.878	2.348
84	3.6	5.851	2.251
87	3.6	5.768	2.168
88	3.56	5.908	2.348
89	3.6	5.925	2.325
92	3.51	5.864	2.354
93	5.18	5.893	0.713
94	4.74	5.211	0.471
95	4.68	5.938	1.258
97	5.47	6.013	0.543
98	3.6	5.935	2.335
99	4.92	5.803	0.883
100	4.89	5.712	0.822
101	4.72	5.994	1.274
102	4	5.839	1.839
103	4.52	5.363	0.843
105	4.7	5.904	1.204
106	3.82	5.9	2.08
107	3.82	5.896	2.076

Tableau III.13 Les valeurs expérimentales, prédites et résiduelles des activités par MLP sur les données de test.

N°	Observée	Calculée	Résiduel
4	5.59	5.925	0.335
6	6.92	5.88	-1.04
10	5.24	5.85	0.61
19	5.14	5.808	0.668
20	5	5.653	0.653
24	7.23	8.088	0.858
29	5.47	5.85	0.38
31	7.89	5.878	-2.012
36	6.01	7.776	1.766
40	4.37	5.847	1.477
55	8.23	5.887	-2.343
57	8.09	8.18	0.09
69	5.79	8.106	2.316
75	6.47	5.859	-0.611
79	7.02	8.167	1.147
85	3.6	5.85	2.25
86	3.72	5.682	1.962
90	3.96	5.782	1.822
91	3.45	5.83	2.38
96	4.74	5.568	0.828
104	4.7	5.823	1.123

III.5 Discussion

La relation d'activité de la structure quantitative est d'un grand intérêt dans la découverte de nouvelles entités chimiques puissantes car elle est utilisée pour prédire les activités de molécules en se basant uniquement sur leurs propriétés chimiques.

Afin d'obtenir de bonnes relations prédictives, il est nécessaire de découvrir et d'utiliser un certain ensemble de descripteurs moléculaires qui ont de bonnes corrélations avec l'activité biologique cible. Pour cela, un nombre d'indicateurs globaux ont été calculé tels que le potentiel chimique, le potentiel d'ionisation, l'affinité électronique, la dureté, la douceur globale, l'électronégativité et l'indice électrophile, ainsi que les indices de Fukui.

Ensuite, des modèles basés sur les méthodes statistiques : régression linéaire et perceptron multicouche ont été développés pour l'inhibition de la transcriptase inverse du VIH en utilisant des informations issues de la chimie quantique. Les modèles QSAR ont été construits sur un

ensemble d'apprentissage et validé par rapport à un ensemble de test. L'activité prédite a ensuite été comparée à la valeur réelle (expérimentale).

Les résultats montrent que les méthodes utilisées permettent de prédire l'activité étudiée et peuvent être utiles pour prédire l'activité biologique de nouveaux composés. Les validations des modèles ont été utilisées pour déterminer la qualité et la puissance statistique des modèles prédictif MLR et MLP QSAR.

Le modèle basé sur MLR est capable de décrire environ 56.11% de l'activité expérimentale sur les données d'apprentissage et 63.59 % sur les données de test et pourrait être utilisé efficacement pour estimer l'activité des dérivés de HEPT. Notre étude montre aussi que le modèle MLP est également considéré outil efficace pour prédire l'activité inhibitrice dans la maladie du VIH basée sur Descripteurs suggérés avec une corrélation de 58.26% sur les données d'apprentissage et 49.33% sur les données de test.

Les résultats peuvent varier en fonction des types des descripteurs et la méthode statistique utilisé mais cela n'exclut pas que le faite que les indices de Fukui constituent de très bons descripteurs pour les modèles QSAR.

III.6 Conclusion

QSAR est essentiellement une approche statistique corrélant la propriété de réponse ou les données d'activité avec des descripteurs codant des informations chimiques.

Dans ce chapitre, l'activité inhibitrice de la transcriptase inverse du VIH a été modélisée par deux techniques d'apprentissage automatique différentes, à savoir la régression linéaire et le perceptron multicouche sur un ensemble de 107 dérivés de HEPT en utilisant des informations issues de la chimie quantique.

Par conséquent, le modèle QSAR basé sur la régression linéaire peut être considéré comme une approche prometteuse pour la prédiction de l'activité étudiée sur l'ensembles des 107 dérivés HEPT. En outre, le modèle doit être testé avec d'autres de descripteurs pour authentifier sa précision prédictive.

Bibliographie

- [1] TOUHAMI Moufida. 2019. Modélisation de l'activité biologique de composés hétérocycles. Thèse de doctorat. Université de Saida. Algérie.
- [2] Juan M. Luco et Ferdinando H. Ferretti. 1997. QSAR Based on Multiple Linear Regression and PLS Methods for the Anti-HIV Activity of a Large Group of HEPT Derivatives. *Journal of chemical information and computer sciences*, 37(2), 392-401.
- [3] http://teqip.jdvu.ac.in/QSAR_Tools/
- [4] Hansch, C.; Zhang, L. QSAR of HIV Inhibitors. *Bioorg. Med. Chem. Lett.* 1992, 2, 1165-1169.
- [5] Hannongbua, S., Lawtrakul, L., & Limtrakul, J. (1996). Structure-activity correlation study of HIV-1 inhibitors: Electronic and molecular parameters. *Journal of computer aided molecular design*, 10(2), 145-152.
- [6] Kireev, D. B., Chrétien, J. R., Grierson, D. S., & Monneret, C. (1997). A 3D QSAR study of a series of HEPT analogues: the influence of conformational mobility on HIV-1 reverse transcriptase inhibition. *Journal of medicinal chemistry*, 40(26), 4257-4264.
- [7] Jalali-Heravi, M., & Parastar, F. (2000). Use of artificial neural networks in a QSAR study of anti-HIV activity for a large group of HEPT derivatives. *Journal of chemical information and computer sciences*, 40(1), 147-154.
- [8] Bazoui, H., Zahouily, M., Boulajaaj, S., Sebti, S., & Zakarya, D. (2002). QSAR for anti-HIV activity of HEPT derivatives. *SAR and QSAR in Environmental Research*, 13(6), 567-577.
- [9] Douali, L., Villemin, D., Zyad, A., & Cherqaoui, D. (2004). Artificial neural networks: Non-linear QSAR studies of HEPT derivatives as HIV-1 reverse transcriptase inhibitors. *Molecular diversity*, 8(1), 1.
- [10] Prakasvudhisarn, C., & Lawtrakul, L. (2008). Feature set selection in QSAR of 1-[(2-Hydroxyethoxy) methyl]-6-(phenylthio) thymine (HEPT) analogues by using swarm intelligence. *Monatshefte für Chemie-Chemical Monthly*, 139(3), 197-211.
- [11] Darnag, R., Schmitzer, A., Belmiloud, Y., Villemin, D., Jarid, A., Chait, A., ... & Cherqaoui, D. (2009). QSAR studies of HEPT derivatives using support vector machines. *Molecular Informatics*, 28(6-7), 709-718.
- [12] Inthajak, K., Khamsemanan, N., Nattee, C., Toochinda, P., & Lawtrakul, L. (2017). A prediction approach for anti-HIV activity of HEPT compounds using random forest technique. *Monatshefte für Chemie-Chemical Monthly*, 148(10), 1697-1709.
- [13] Tomar, S., Nadreen, B., Kaleem, A. (2015). QSAR study of anti-HIV drugs of 1-2-

[(hydroxyethoxy) methyl]-6-(phenylthio) thymine (HEPT) derivatives by using quantum descriptors. *Journal of Chemical and Pharmaceutical Research*, 7(3), 146-151.

[14] Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. 2016. WEKA.

Conclusion Générale

Conclusion Générale

La relation quantitative structure-activité (QSAR) est un processus dans lequel une structure chimique est quantitativement liée à un processus bien défini (tel qu'une activité biologique ou une réactivité chimique).

Les relations quantitatives structure-activité sont utilisées pour prédire les activités de molécules en se basant uniquement sur leurs structures chimiques, pour construire des modèles fiables, robustes, stables et précis, capables de prédire efficacement cette activité.

Dans ce travail, une étude quantitative de la relation structure-activité (QSAR) a été effectuée pour déterminer quantitativement l'effet de la structure moléculaire des composés étudiés sur leur activité biologique. Des modèles mathématiques ont été développés pour prédire l'activité biologique inhibitrice des dérivés HEPT.

Les descripteurs utilisés dans cette étude ont été calculés par la méthode DFT avec la fonction B3LYP associée à un ensemble de base 6-311G. Pour cela nous avons, dans un premier temps, mené notre étude à partir de la synthèse et le développement de la molécule HEPT et ses dérivés. Des calculs de chimie quantique ont été appliqués sur l'ensemble de molécules dérivées de HEPT. Nous nous intéressons à l'étude de leurs caractéristiques structurales et de leurs propriétés électroniques. En nous appuyant sur l'énergie du système, nous pouvons identifier divers indicateurs globaux tels que le potentiel chimique, le potentiel d'ionisation, l'affinité électronique, la dureté, la douceur globale, l'électronégativité et l'indice électrophile. D'autre part, en utilisant la théorie de l'analyse de la population, l'indice de Fukui a été calculé.

De nombreuses méthodes statistiques peuvent être mises en œuvre pour la construction de modèles QSAR, parmi les plus accessibles : les régressions linéaires multiples (Multiple Linear Regression ou MLR), et perceptron multi couche (MLP) qui sont considérés comme des outils efficaces pour prédire l'activité inhibitrice de plusieurs maladies tels que le virus de l'immunodéficience humaine.

La validité des modèles obtenus a confirmé le pouvoir prédictif des modèles obtenus. Les valeurs prédites sont plus proches que les valeurs expérimentales pour l'activité biologique.

Les perspectives de ce travail sont multiples. D'une part l'intégration d'autres descripteurs sera bénéfique pour améliorer la prédictivité des modèles obtenus, d'autre part, un prolongement évident dans la modélisation qualitative et précisément avec les méthodes de régression et de classification comme l'analyse par composantes principales ACP, machine à support vecteur SVM, ...etc.